

# Enjeux et défis des techniques de compression pour les données de simulation massives

## CORESA 2021

---

Laurent Duval

*with:* Lauriane Bouard, Frédéric Payan, Marc Antonini, Christophe Preux

3 novembre 2021

Context: multimedia versus simulation data, what is different?

Issues in simulation: example from geoscience

Toward solutions: state-of-the-art

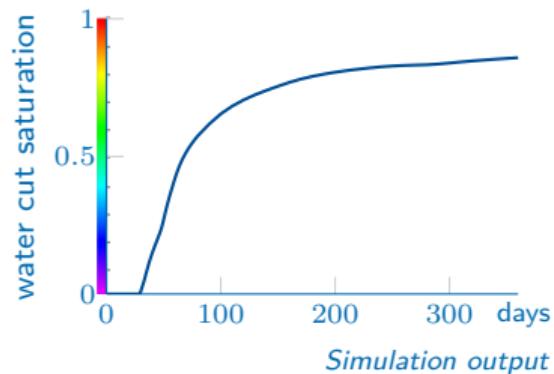
Workflow: embedded with compression

Compression: methodology and steps

Compression: objective and subjective results

Simulation: subjective results and better adapted metrics

Challenges and opportunities



Compression in simulation: what for?

*Flow simulation in a reservoir*

A pervasive and transparent paradigm

A world of streaming and evanescent binary abstractions (Deeflix, Netzer, SpotS, OCify)

Three most common types of objects

Audio (essentially 1D)

Image (essentially 2D)

Video (essentially 2D+1D)

Mostly meant for (human) interpretation

Knowledge of the human sensory system

Wide audience for testing ("all users")

Little need for post-processing\*

Reasonably well-structured data

Audio: even sampling

Image: cartesian pixel grid (mostly)

Video: audio+image at regular frame rates (24/25 fps)

## On data-intensive scientific discovery

So-called "fourth paradigm" (Hey, Tolle, Tansley, 2009)

Three different ones: empirical, theoretical, computational science (no drama)

Tentative layers of "scientific data" object abstraction:

Instrumental (sensors, recording devices, IoT)

Computational (modeling, simulation, avatar, metaverse)

Inferential (statistical model, data science, machine/deep learning)

Mostly meant for (reality) understanding

Often outside human sight range

Sharing in relatively small scientific communities

Important postprocessing or workflows for further data exploitation

Unreasonably ill-structured data

Many in-house or ad-hoc storage structures (e.g. HDF, CDF)

Not sufficiently used for *imposing* pervasive "formats" (nor well documented)

*The nice thing about standards is that you have so many to choose from; furthermore, if you do not like any of them, you can just wait for next year's model.*" (Attributed to A. S. Tanenbaum)

What makes sense can be shrunk: redundancy versus relevancy

- Part is predictable (**resolution**)
- Part is forgetable (**precision**)

Main actions

- Transform data to enhance the useful (cosine, wavelet, companding)
- Discard (or reorder) information from useful to less useful
- Encode binary streams in sensible ways

Goals

- Lossless, lossy
- Quasi-lossless, lossless to lossy, progressive (in resolution or precision)
- Efficiency, usability

## Multimedia data

Necessary standardization (mp3: 1991; jpeg: 1992; mpeg4: 1998\*)

Small-to-medium size (fits on a "drive")

Many objective/subjective models and metrics

In one equation:  $I - \hat{I} = i$  (integer)

## Data-intensive science

Storage and HPC inflation

Bigger size (sometimes fit on several drives\*)

Way less objective/subjective models and metrics

In a second equation:  $4F - \widehat{4F} = 4F$  (quadruple-precision floating-point)

Geophysics in 2012: 2-3 months to sort (byte-swapping\*)

Computations to train largest NLP models: increased 300,000 times in 6 years (April 2021)

GPT-4: 100 trillion parameters, "five hundred times" larger than GPT-3

Microscopy (TEM, SEM): 50 Tb/sample, thousands of cell slides (biology\*)

CERN Large Hadron Collider: 90+25 petabytes/year (physics)

Climatology: "[Coupled Model Intercomparison Project (CIMP6)] entire project is estimated to release about 20 to 40 petabytes of data from more than 20 climate models"

Quite often, massive simulations produce

more data than we can store\*

more quickly than we can write

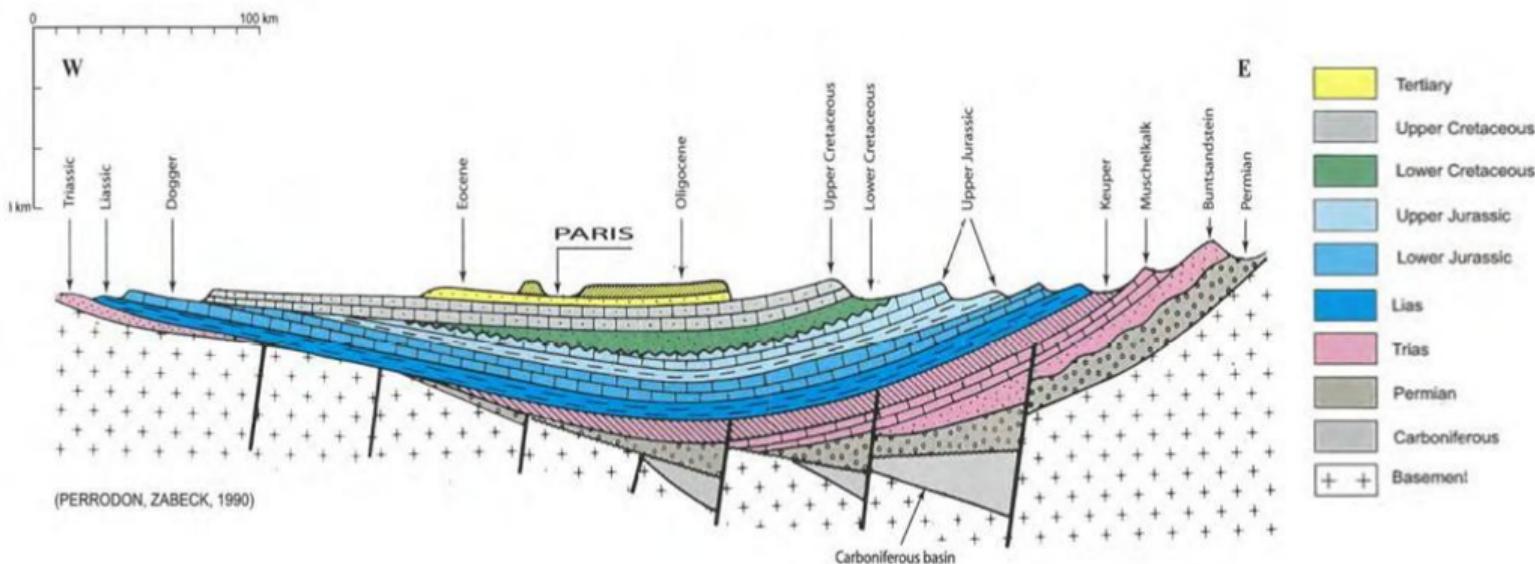
more data than they can share\*

more data than we can pay for (material, energetic, legal)

# Geological data

## Section of sedimentary basin

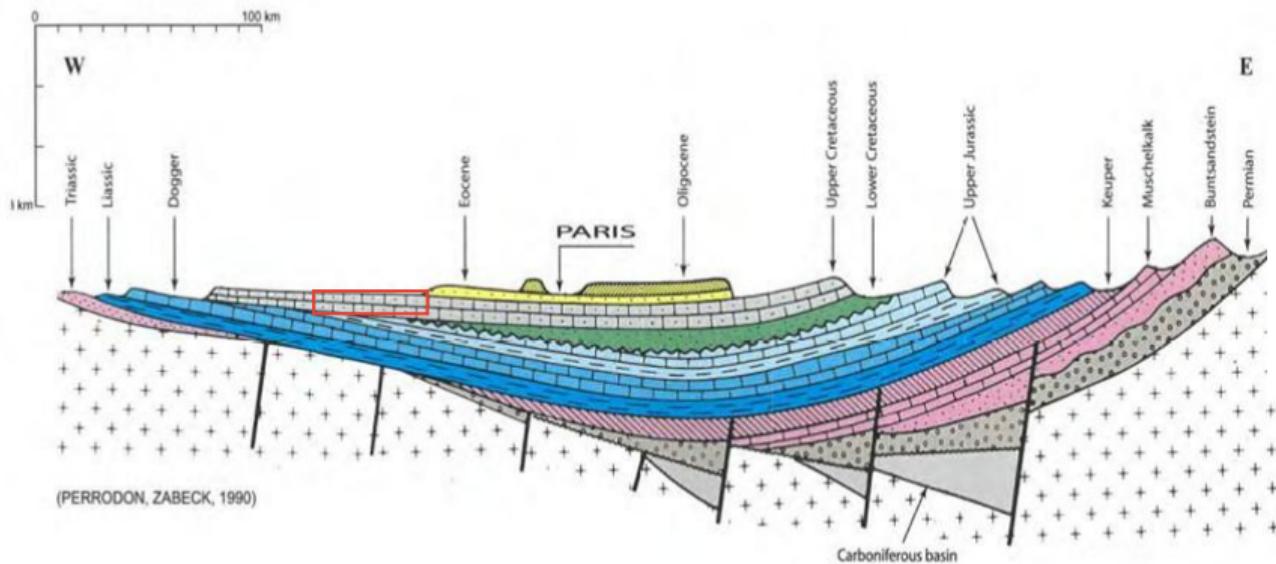
## Issues in simulation



# Geological data

## Section of sedimentary basin

## Issues in simulation



### Structure: Corner Point Grid format (CPG)

Pillar extremities

2D

$(x, y, z)$  coordinates

Node altitudes

1D

relative

Cell activity

3D

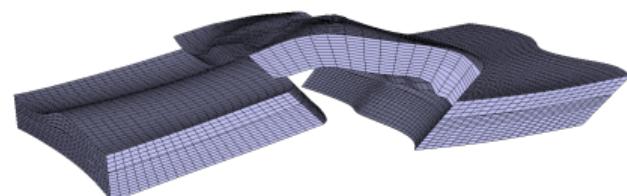
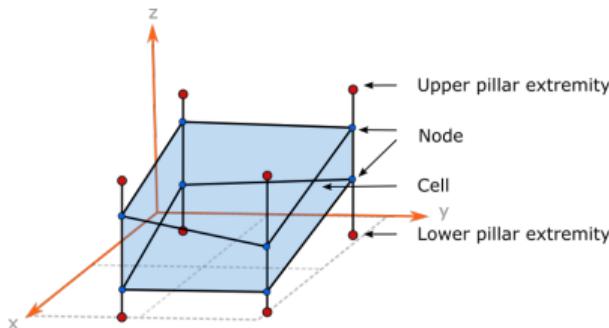
(boolean/cell)

Properties

3D

Categorical (low dynamic)

(discrete-integer/cell)



### Structure: Corner Point Grid format (CPG)

Pillar extremities                     $2D$

Node altitudes                     $1D$

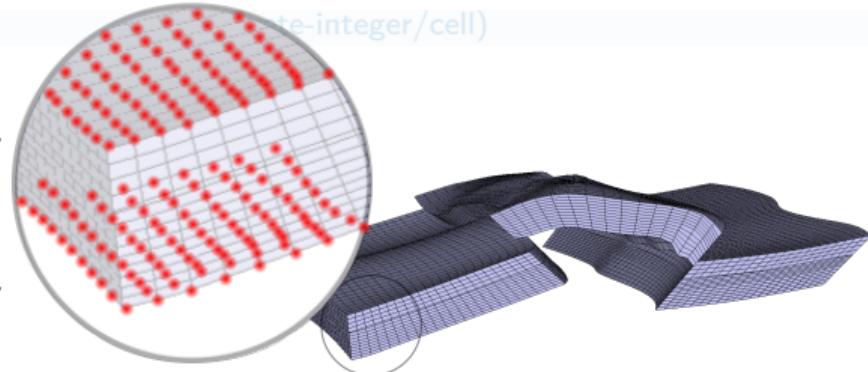
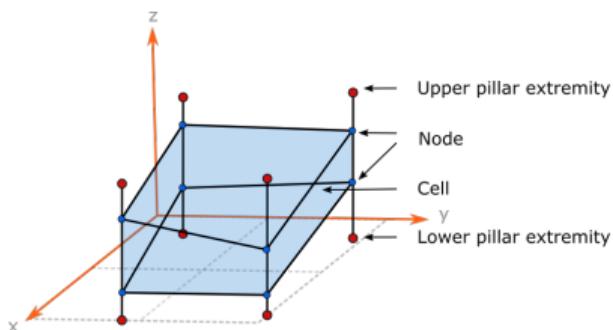
Cell activity                     $3D$

Properties                     $3D$

$(x, y, z)$  coordinates

relative  
(boolean/cell)

### Categorical (low dynamic)



### Structure: Corner Point Grid format (CPG)

Pillar extremities       $2D$

Node altitudes       $1D$

$(x, y, z)$  coordinates

relative

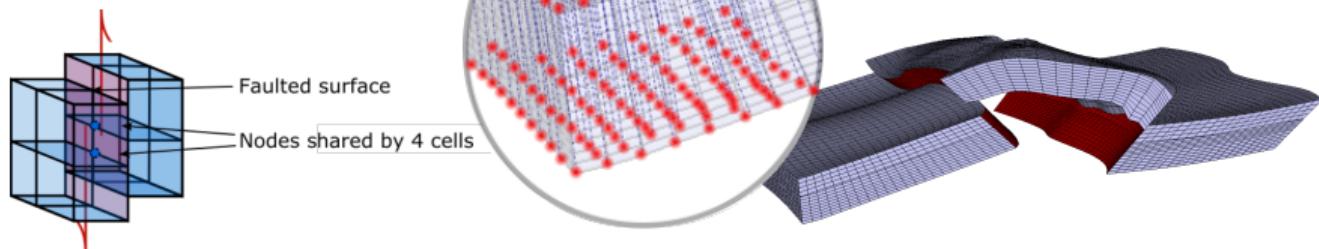
Cell activity       $3D$

(boolean/cell)

Properties       $3D$

Categorical (low dynamic)

Finite-integer/cell)



# Geological data model

## Issues in simulation

### Structure: Corner Point Grid format (CPG)

Pillar extremities                     $2D$

Node altitudes                     $1D$

Cell activity                         $3D$

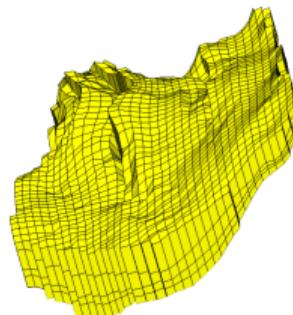
Properties                          $3D$

$(x, y, z)$  coordinates

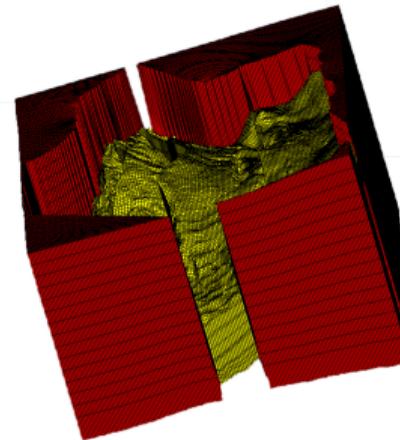
relative

(boolean/cell)

Categorical (low dynamic)



(different interior/color)



# Geological data model

Issues in simulation

## Structure: Corner Point Grid format (CPG)

Pillar extremities       $2D$        $(x, y, z)$  coordinates

Node altitudes       $1D$       relative

Cell activity       $3D$       (boolean/cell)

Properties ( $\mathbb{R}$ )       $3D$       (floating-point/cell)

Categorical (low dynamic)      (discrete-integer/cell)

# Geological data model

## Issues in simulation

Structure: Corner Point Grid format (CPG)

Pillar extremities       $2D$

Node altitudes       $1D$

Cell activity       $3D$

**Properties ( $\mathbb{R}$ )**       $3D$

Porosity (high dynamic)

Categorical (low dynamic)

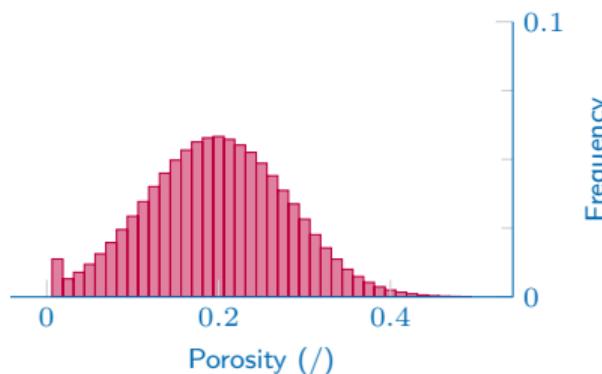
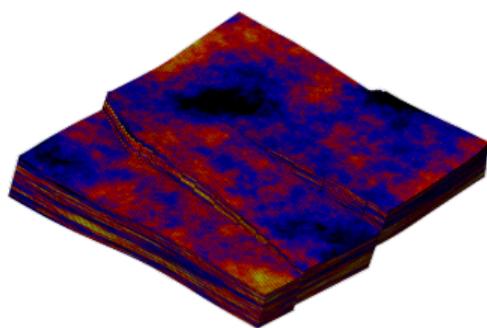
$(x, y, z)$  coordinates

relative

(boolean/cell)

(floating-point/cell)

(discrete-integer/cell)



# Geological data model

## Issues in simulation

Structure: Corner Point Grid format (CPG)

Pillar extremities       $2D$

Node altitudes       $1D$

Cell activity       $3D$

Properties ( $\mathbb{R}$ )       $3D$

Permeability (high dynamic)

Categorical (low dynamic)

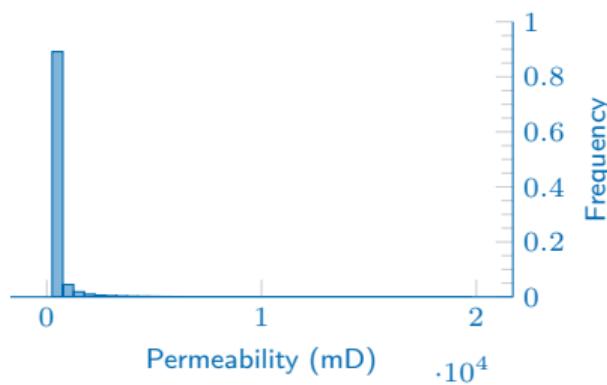
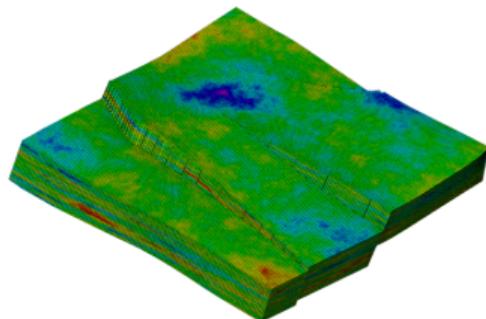
$(x, y, z)$  coordinates

relative

(boolean/cell)

(floating-point/cell)

(discrete-integer/cell)



# Geological data model

Issues in simulation

Structure: Corner Point Grid format (CPG)

Pillar extremities       $2D$

$(x, y, z)$  coordinates

Node altitudes       $1D$

relative

Cell activity       $3D$

(boolean/cell)

Properties ( $\mathbb{N}$ )       $3D$

Categorical (low dynamic)

(discrete-integer/cell)

Issues, compressed

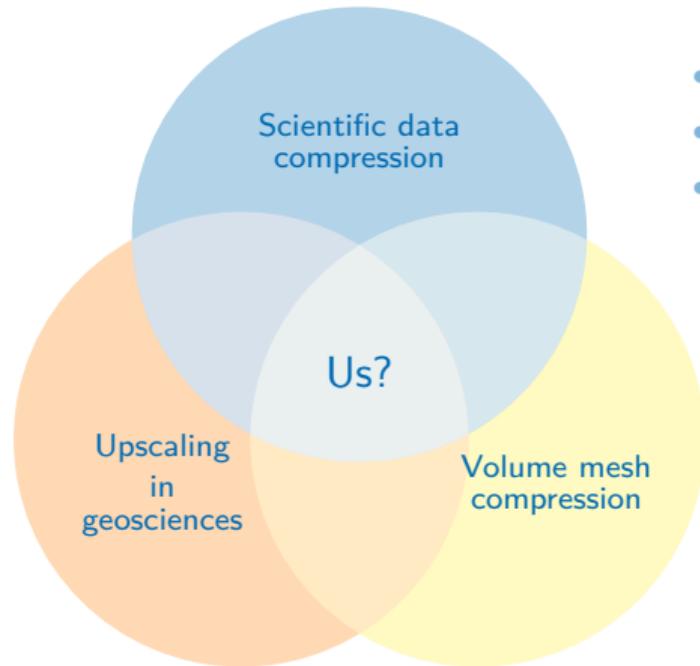
Issues in simulation

72 hours simulation on an average scientific laptop

# A compact state-of-the-start (them and us)

Toward solutions

- Durlofsky, 2012
- Li, 2000
- Qi, 2009



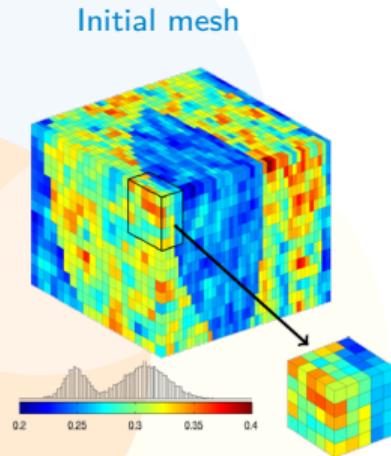
- Cappello, 2019
- Lindstrom, 2014
- Baker, 2019

# A compact state-of-the-start (them and us)

Toward solutions

- Durlofsky, 2012
- Li, 2000
- Qi, 2009

Upscaling  
in  
geosciences



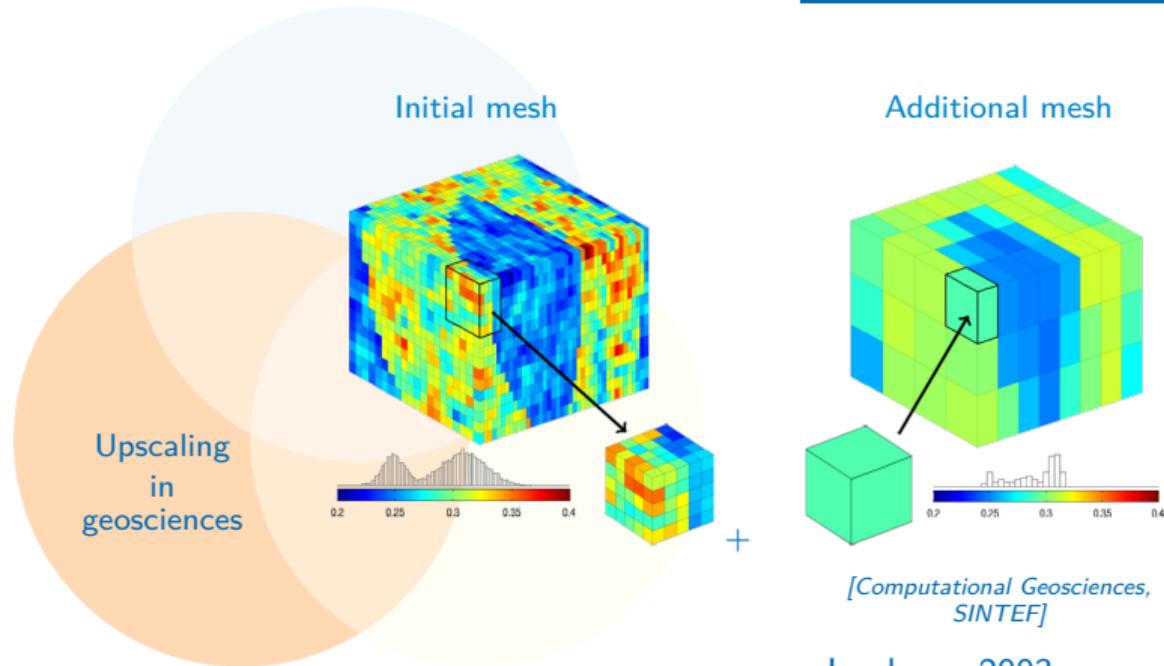
- Developed for simulation
- Not for storage

# A compact state-of-the-start (them and us)

Toward solutions

- Durlofsky, 2012
- Li, 2000
- Qi, 2009

- Developed for simulation
- Not for storage



- Isenburg, 2003
- Krivograd, 2008
- Lindstrom, 2008

# A compact state-of-the-start (them and us)

Toward solutions

- Cappello, 2019
- Lindstrom, 2014
- Baker, 2019

The diagram features a large light blue circle labeled "Scientific data compression". Inside it, there is a smaller light orange circle labeled "Transformation". Below these circles, a yellow circle overlaps them and is labeled "Refinable precision". The overlapping areas between the circles are shaded in a darker shade of their respective colors.

	Preprocess	Blocking	Transformation	Quantization	Encoding
SZ	float to int	flatten array	prediction	uniform	Huffman
ZFP		subblock	fixed mantissa/subblock, DCT		Zerotree

*Refinable precision*

# A compact state-of-the-start (them and us)

Toward solutions

- Cappello, 2019
- Lindstrom, 2014
- Baker, 2019

Scientific data compression

	Preprocess	Blocking	Transformation	Quantization	Encoding
SZ	float to int	flatten array	prediction	uniform	Huffman
ZFP		subblock	fixed mantissa/subblock, DCT		Zerotree
HexaShrink	float to int		Dyadic wavelet		Zerotree

*Refinable precision*

*Refinable resolution & precision*

# A compact state-of-the-start (them and us)

Toward solutions

## Scientific data compression

- Cappello, 2019
- Lindstrom, 2014
- Baker, 2019

	Preprocess	Blocking	Transformation	Quantization	Encoding
SZ	float to int	flatten array	prediction	uniform	Huffman
ZFP		subblock	fixed mantissa/subblock, DCT		Zerotree

*Refinable precision*

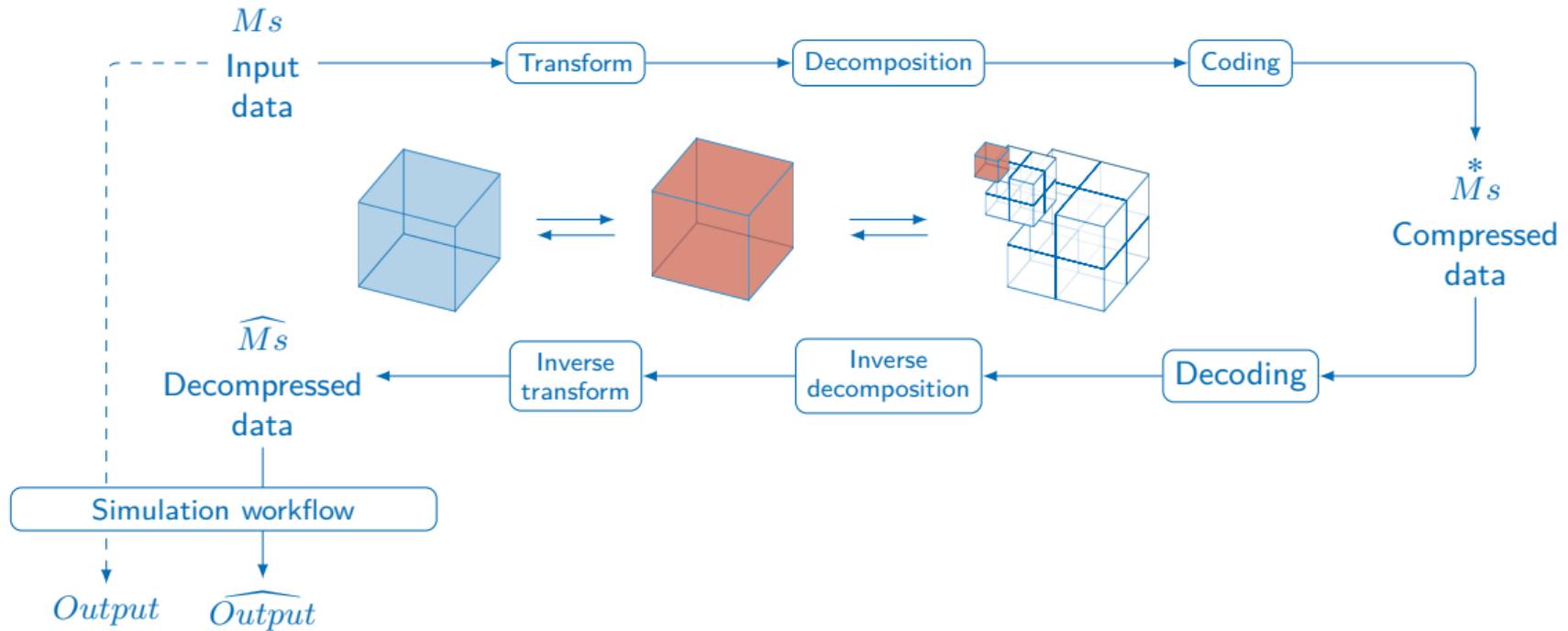
HexaShrink

float to int		Dyadic wavelet	Zerotree
--------------	--	----------------	----------

*Refinable resolution & precision*

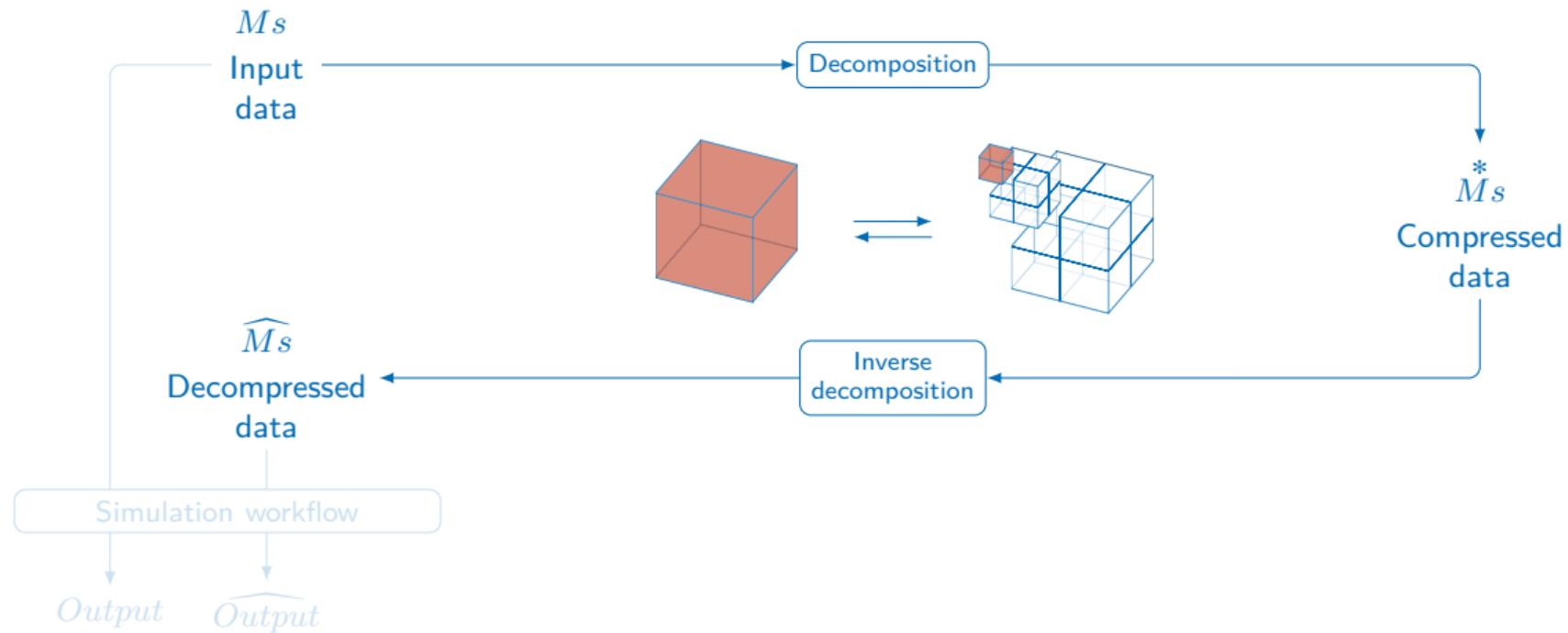
## Generic simulation workflow

Workflow



# HexaShrink (HS) multiresolution decomposition

Workflow



# HexaShrink (HS) multiresolution decomposition

Workflow

## Structure

Pillar extremities  
Node altitudes

Surface points  
1D data

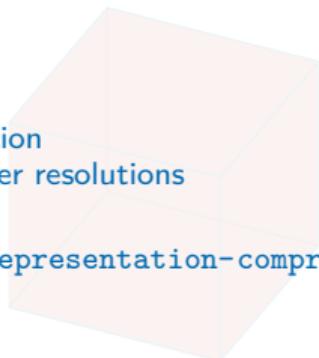
2D morphological wavelet  
Non linear 1D wavelet

## Cell activity Properties

3D data (boolean)  
3D data

3D wavelet

Initial data



## Motivations:

- Structure & faults preservation
- Properties coherency at lower resolutions

[http://www.laurent-duval.eu/  
opus-hexashrink-multiscale-mesh-representation-compression-wavelet-modelet.  
html](http://www.laurent-duval.eu/opus-hexashrink-multiscale-mesh-representation-compression-wavelet-modelet.html)



Initial mesh

# HexaShrink (HS) multiresolution decomposition

Workflow

## Structure

Pillar extremities  
Node altitudes

Surface points  
1D data

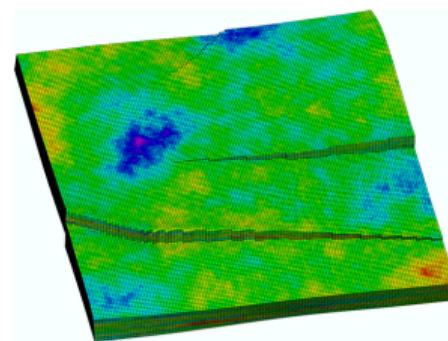
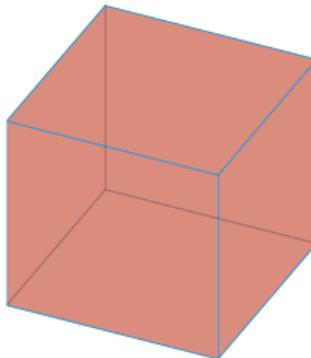
2D morphological wavelet  
Non linear 1D wavelet

Cell activity  
Properties

3D data (boolean)  
3D data

3D wavelet

Initial data



Initial mesh

# HexaShrink (HS) multiresolution decomposition

Workflow

## Structure

Pillar extremities  
Node altitudes

Surface points  
1D data

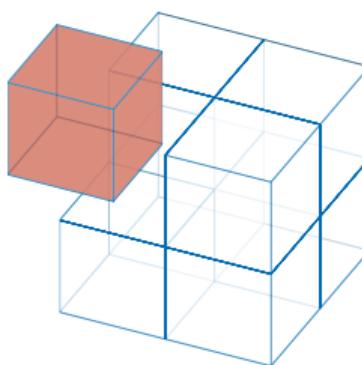
2D morphological wavelet  
Non linear 1D wavelet

Cell activity  
Properties

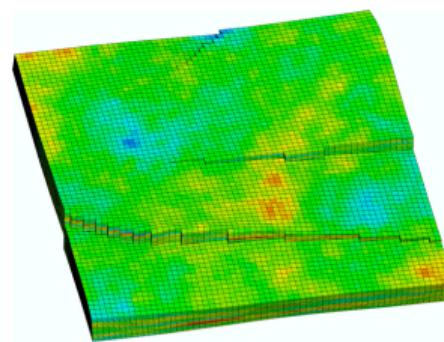
3D data (boolean)  
3D data

3D wavelet

An.



$D^{-1}$



Generated from Analysis (An.) output

# HexaShrink (HS) multiresolution decomposition

Workflow

## Structure

Pillar extremities  
Node altitudes

Surface points  
1D data

2D morphological wavelet  
Non linear 1D wavelet

Cell activity  
Properties

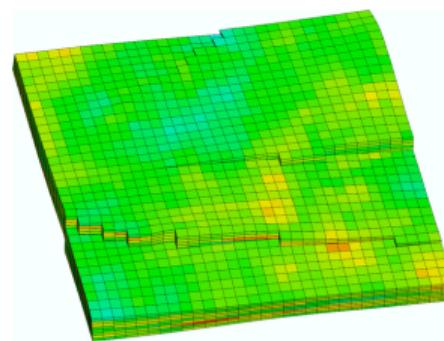
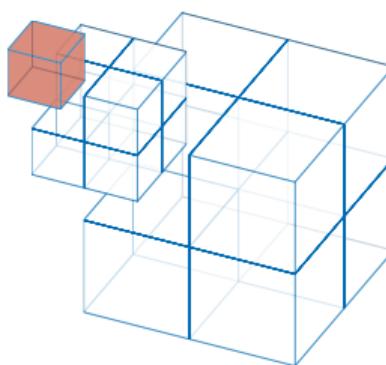
3D data (boolean)  
3D data

3D wavelet

An.

$D^{-2}$

$D^{-1}$



Generated from Analysis (An.) output

# HexaShrink (HS) multiresolution decomposition

Workflow

## Structure

Pillar extremities  
Node altitudes

Surface points  
1D data

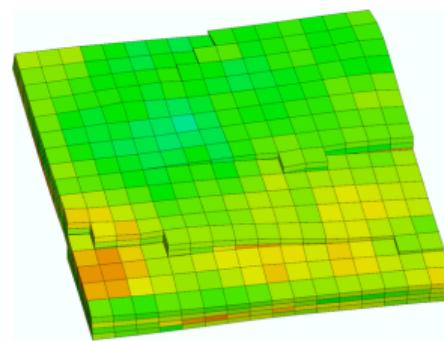
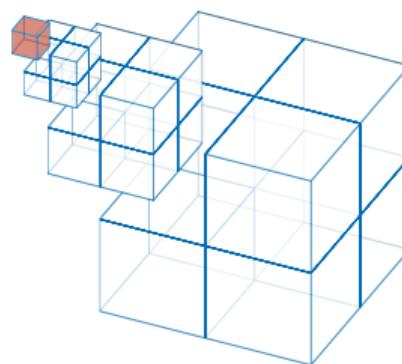
2D morphological wavelet  
Non linear 1D wavelet

Cell activity  
Properties

3D data (boolean)  
3D data

3D wavelet

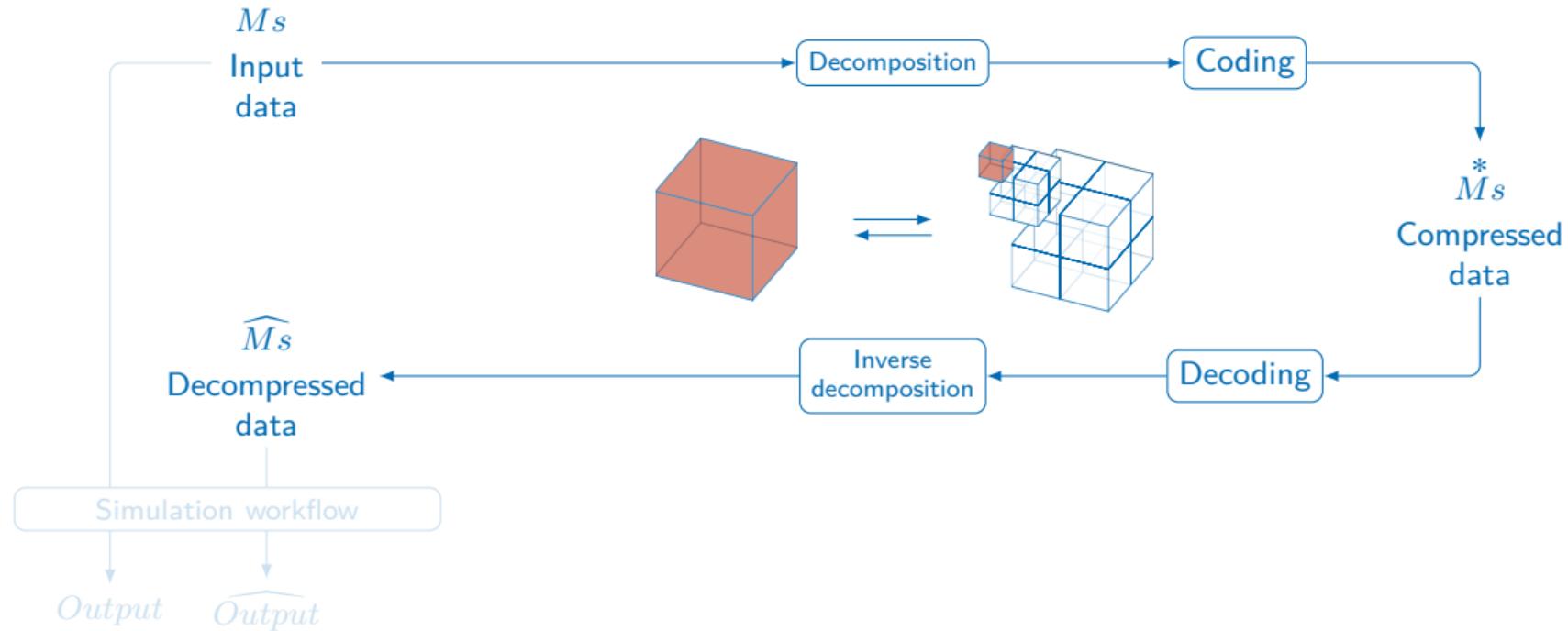
An.  
 $D^{-3}$   
 $D^{-2}$   
 $D^{-1}$



Generated from Analysis (An.) output

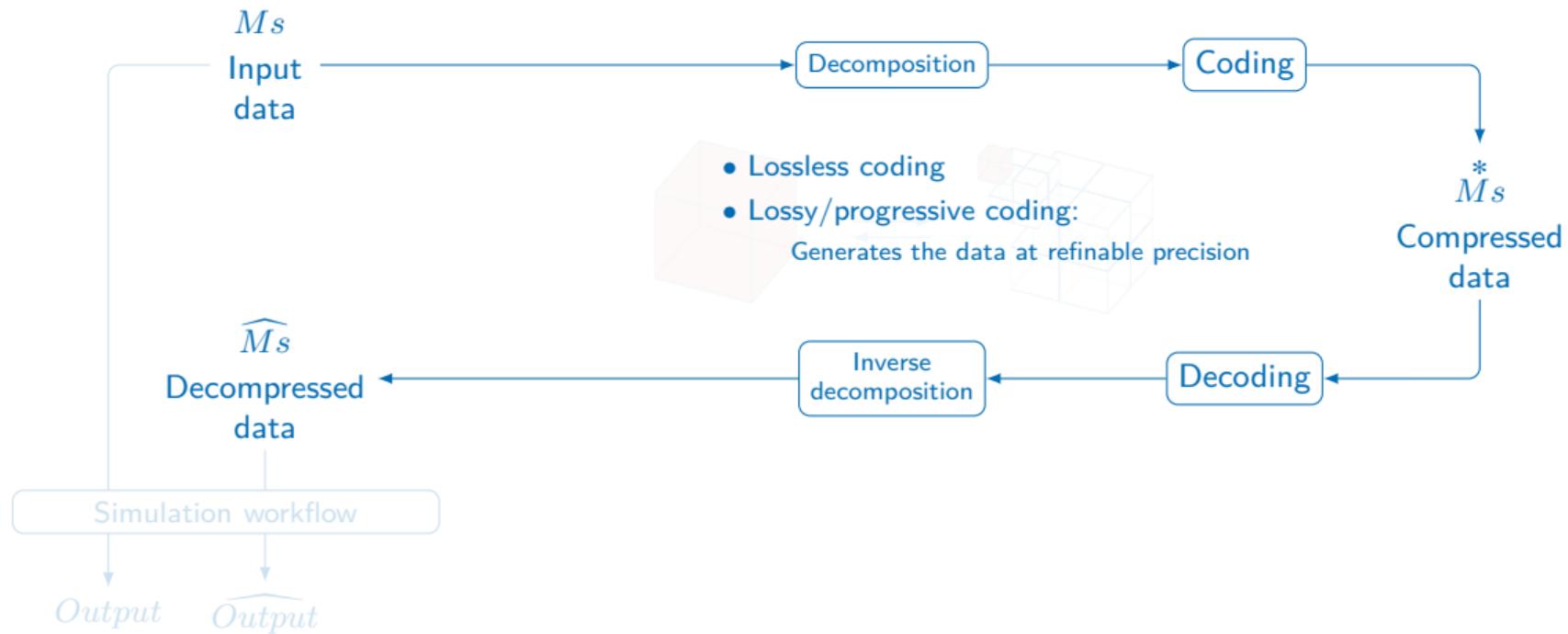
# HexaShrink coding

Compression steps



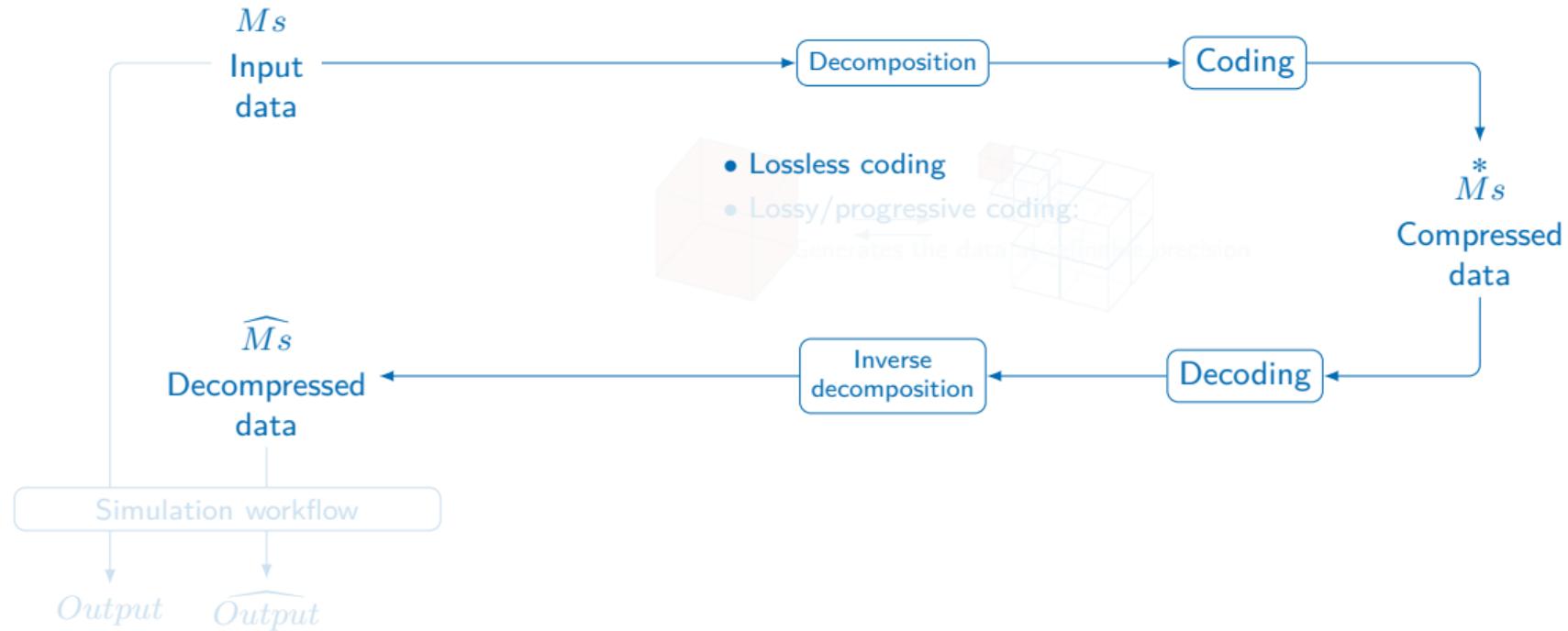
# HexaShrink coding

Compression steps



# HexaShrink coding

Compression steps



# HexaShrink lossless

Compression steps

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Original data

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Original data

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

4	13	21	28	-1	-9	-2	-4
5	22	22	16	-5	2	1	-1
10	6	12	18	-1	-2	2	8
14	10	6	15	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

1 decomposition level

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

4	13	21	28	-1	-9	-2	-4
5	22	22	16	-5	2	1	-1
10	6	12	18	-1	-2	2	8
14	10	6	15	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

1 decomposition level

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

11	22	-13	-1	-1	-9	-2	-4
10	13	4	-7	-5	2	1	-1
-5	6	9	-12	-1	-2	2	8
-5	5	-1	4	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

2 decomposition levels

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

11	22	-13	-1	-1	-9	-2	-4
10	13	4	-7	-5	2	1	-1
-5	6	9	-12	-1	-2	2	8
-5	5	-1	4	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

Res. -2

11	22
10	13

2 decomposition levels

# HexaShrink lossless

Compression steps

Res. 0

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

11	22	-13	-1	-1	-9	-2	-4
10	13	4	-7	-5	2	1	-1
-5	6	9	-12	-1	-2	2	8
-5	5	-1	4	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

Res. -2

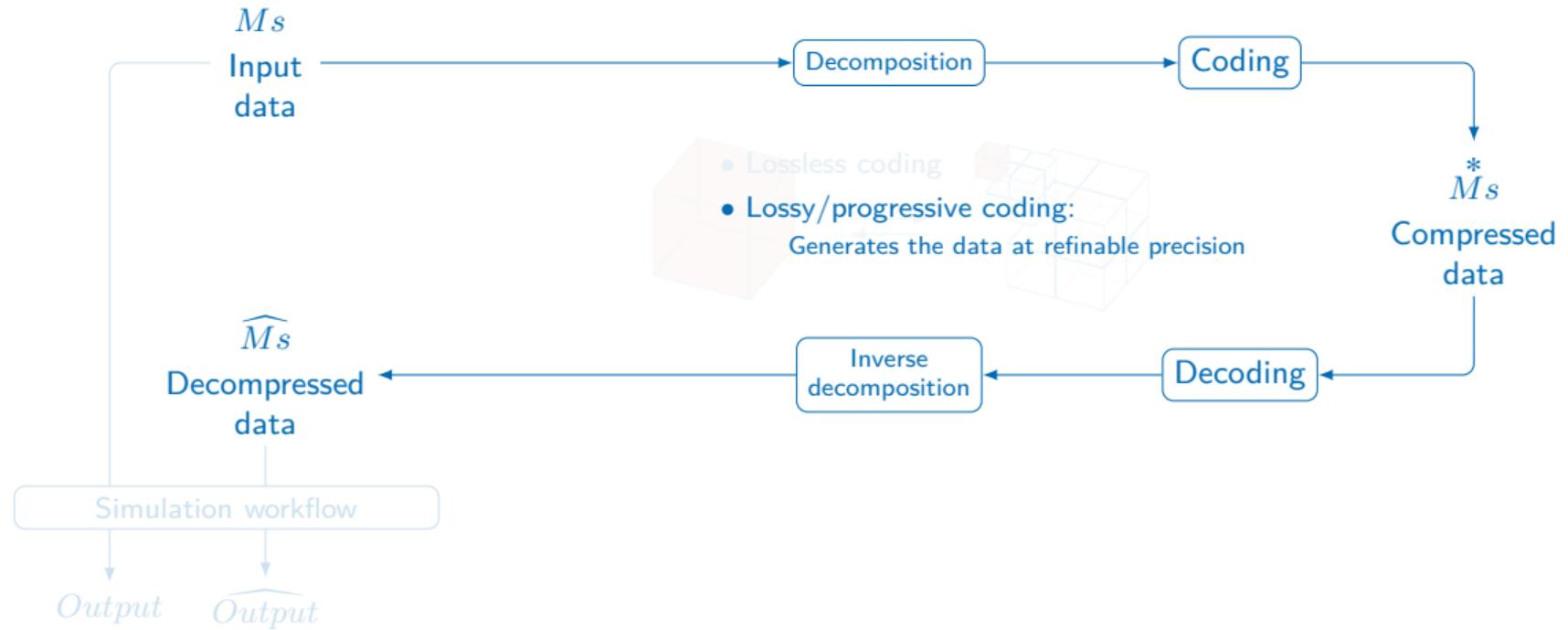
11	22
10	13

2 decomposition levels

Subbands coding  
with generic lossless methods

# HexaShrink lossy

Compression steps



# HexaShrink lossy/progressive

Compression steps

5 MSB<sub>Z</sub>

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. 0

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

Res. -1

11	22	-13	-1	-1	-9	-2	-4
10	13	4	-7	-5	2	1	-1
-5	6	9	-12	-1	-2	2	8
-5	5	-1	4	-1	2	-1	-16
-2	-6	3	3	-5	9	-6	-1
-6	1	3	3	5	7	2	0
-6	-7	16	-2	8	2	5	1
-2	-6	0	-5	7	-5	-4	1

Res. -2

11	22
10	13

Most Significant Bit

4	11
3	0
2	1
1	0
0	1

5 MSB<sub>Z</sub>

Least Significant Bit

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>							
2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

10	22	-12	0	0	-8	-2	-4
10	12	4	-6	-4	2	0	0
-4	6	8	-12	0	-2	2	8
-4	4	0	4	0	2	0	-16
-2	-6	2	2	-4	8	-6	0
-6	0	2	2	4	6	2	0
-6	-6	16	-2	8	2	4	0
-2	-6	0	-4	6	-4	-4	0

Res. -2

11	22
10	13

Most Significant Bit

4  
3  
2  
1  
0

4 MSB<sub>Z</sub>

2  
1  
0

Least Significant Bit

0



# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>							
2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

4 MSB <sub>Z</sub>							
0	4	6	12	16	21	25	30
4	3	8	22	17	16	24	27
0	3	24	21	22	21	16	18
3	11	21	24	19	20	14	16
6	4	2	3	19	17	20	14
8	14	9	12	3	5	22	16
14	13	7	10	4	8	4	20
13	18	16	13	6	6	10	26

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

10	22	-12	0	0	-8	-2	-4
10	12	4	-6	-4	2	0	0
-4	6	8	-12	0	-2	2	8
-4	4	0	4	0	2	0	-16
-2	-6	2	2	-4	8	-6	0
-6	0	2	2	4	6	2	0
-6	-6	16	-2	8	2	4	0
-2	-6	0	-4	6	-4	-4	0

Res. -2

11	22
10	13

Most Significant Bit

4

0

1

4 MSB<sub>Z</sub>

2

0

1

Least Significant Bit

0

1

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>							
2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

4 MSB <sub>Z</sub>							
0	4	6	12	16	21	25	30
4	3	8	22	17	16	24	27
0	3	24	21	22	21	16	18
3	11	21	24	19	20	14	16
6	4	2	3	19	17	20	14
8	14	9	12	3	5	22	16
14	13	7	10	4	8	4	20
13	18	16	13	6	6	10	26

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

8	20	-12	0	0	-8	0	-4
8	12	4	-4	-4	0	0	0
-4	4	8	-12	0	0	0	8
-4	4	0	4	0	0	0	-16
0	-4	0	0	-4	8	-4	0
-4	0	0	0	4	4	0	0
-4	-4	16	0	8	0	4	0
0	-4	0	-4	4	-4	-4	0

Res. -2

11	22
10	13

Most Significant Bit

8

0

3

1

3 MSB<sub>Z</sub>

0

2

1

Least Significant Bit

0

1

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>				
2	5	8	12	20
6	4	9	22	20
1	3	25	20	24
4	11	21	23	20
8	5	2	3	22
10	15	8	11	4
15	12	7	8	4
13	17	15	11	6
5	9	25		

4 MSB <sub>Z</sub>				
0	4	6	12	16
4	3	8	22	17
0	3	24	21	22
3	11	21	24	19
6	4	2	3	19
8	14	9	12	3
14	13	7	10	4
13	18	16	13	6
10	6	10	10	26

3 MSB <sub>Z</sub>				
0	0	2	10	12
0	0	6	22	16
0	1	23	21	20
1	11	21	23	20
0	0	0	3	17
3	11	7	11	3
10	12	5	9	3
12	18	17	13	5
7	10	26		

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

8	20	-12	0	0	-8	0	-4
8	12	4	-4	-4	0	0	0
-4	4	8	-12	0	0	0	8
-4	4	0	4	0	0	0	-16
0	-4	0	0	-4	8	-4	0

Most Significant Bit

4

8
0
1

3 MSB<sub>Z</sub>

3

0
1
0

Least Significant Bit

2

1
0

Res. -2

11	22
10	13

0	-4	0	0	-4	8	-4	0
-4	0	0	0	4	4	0	0
-4	-4	16	0	8	0	4	0
0	-4	0	-4	4	-4	-4	0

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>				
2	5	8	12	20
6	4	9	22	20
1	3	25	20	24
4	11	21	23	20
8	5	2	3	22
10	15	8	11	4
15	12	7	8	4
13	17	15	11	6
5	9	25		

4 MSB <sub>Z</sub>				
0	4	6	12	16
4	3	8	22	17
0	3	24	21	22
3	11	21	24	19
6	4	2	3	19
8	14	9	12	3
14	13	7	10	4
13	18	16	13	6
10	6	10	17	20
26				

3 MSB <sub>Z</sub>				
0	0	2	10	12
0	0	6	22	16
0	1	23	21	20
1	11	21	23	20
0	0	0	3	17
3	11	7	11	3
10	12	5	9	3
12	18	17	13	5
7	10	26		

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

8	16	-8	0	0	-8	0	0
8	8	0	0	0	0	0	0
0	0	8	-8	0	0	0	8
0	0	0	0	0	0	0	-16
0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	16	0	8	0	0	0
0	0	0	0	0	0	0	0

Res. -2

11	22
10	13

Most Significant Bit

4

0

1

0

1

2 MSB<sub>Z</sub>

2

1

1

0

Least Significant Bit

0

1

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>				
2	5	8	12	20
6	4	9	22	20
1	3	25	20	24
4	11	21	23	20
8	5	2	3	22
10	15	8	11	4
15	12	7	8	4
13	17	15	11	6
5	9	2	5	22
13	17	15	11	6
25	9	2	5	22

4 MSB <sub>Z</sub>				
0	4	6	12	25
4	3	8	22	17
0	3	24	21	22
3	11	21	24	19
6	4	2	3	19
8	14	9	12	3
14	13	7	10	4
13	18	16	13	6
10	26	6	10	26

3 MSB <sub>Z</sub>				
0	0	2	10	12
0	0	6	22	16
0	1	23	21	20
1	11	21	23	20
0	0	0	3	17
3	11	7	11	3
10	12	5	9	3
12	18	17	13	5
7	10	26	10	26

2 MSB <sub>Z</sub>				
0	0	0	12	2
0	4	4	24	6
0	2	22	22	16
2	10	22	22	16
0	0	0	6	8
0	12	6	14	0
6	14	4	8	0
14	22	16	12	2
6	8	8	8	24

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

8	16	-8	0	0	-8	0	0
8	8	0	0	0	0	0	0
0	0	8	-8	0	0	0	8
0	0	0	0	0	0	0	-16
0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Res. -2

11	22
10	13

Most Significant Bit

4

3

2

1

0

2 MSB<sub>Z</sub>

Least Significant Bit

0

8

0

1

0

1

0

## HexaShrink lossy/progressive

## Compression steps

5 MSB<sub>Z</sub>

2	5	8	12	20	25	27	31
6	4	9	22	20	19	25	28
1	3	25	20	24	22	17	18
4	11	21	23	20	20	14	15
8	5	2	3	22	18	21	13
10	15	8	11	4	5	22	15
15	12	7	8	4	7	5	20
13	17	15	11	6	5	9	25

4 MSB<sub>Z</sub>

0	4	6	12	16	21	25	30
4	3	8	22	17	16	24	27
0	3	24	21	22	21	16	18
3	11	21	24	19	20	14	16
6	4	2	3	19	17	20	14
8	14	9	12	3	5	22	16
14	13	7	10	4	8	4	20
13	18	16	13	6	6	10	26

3 MSB<sub>Z</sub>

0	0	2	10	12	20	25	31
0	0	6	22	16	16	27	29
0	1	23	21	20	20	14	18
1	11	21	23	20	20	14	18
0	0	0	3	17	15	18	14
3	11	7	11	3	5	22	18
10	12	5	9	3	9	2	18
12	18	17	13	5	7	10	26

2 MSB<sub>Z</sub>

0	0	0	12	2	14	18	30
0	4	4	24	6	10	22	26
0	2	22	22	16	16	12	20
2	10	22	22	16	16	12	20
0	0	0	6	8	8	8	8
0	12	6	14	0	0	16	16
6	14	4	8	0	10	0	16
14	22	16	12	2	6	8	24

Res. 0

Res. -1

Res. -2

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

## Progressive coding with Zerotree:

- EZW (Shapiro, J., 1993)
  - SPIHT (Said, A., Pearlman, W., 1996)

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>				
2	5	8	12	20
6	4	9	22	20
1	3	25	20	24
4	11	21	23	20
8	5	2	3	22
10	15	8	11	4
15	12	7	8	4
13	17	15	11	6
5	9	5	9	25

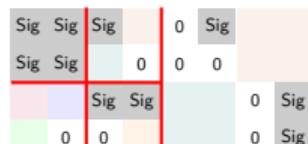
4 MSB <sub>Z</sub>				
0	4	6	12	25
4	3	8	22	17
0	3	24	21	22
3	11	21	24	19
6	4	2	3	19
8	14	9	12	3
14	13	7	10	4
13	18	16	13	6
10	6	10	26	

3 MSB <sub>Z</sub>				
0	0	2	10	12
0	0	6	22	16
0	1	23	21	20
1	11	21	23	20
0	0	0	3	17
3	11	7	11	3
10	12	5	9	3
12	18	17	13	5
7	10	26		

2 MSB <sub>Z</sub>				
0	0	0	12	2
0	4	4	24	6
0	2	22	22	16
2	10	22	22	16
0	0	0	6	8
0	12	6	14	0
6	14	4	8	0
14	22	16	12	2
6	8	8	24	

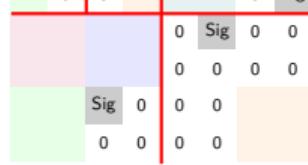
Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15



Res. -2

11	22
10	13



Progressive coding with Zerotree:

- EZW (Shapiro, J., 1993)
- SPIHT (Said, A., Pearlman, W., 1996)

# HexaShrink lossy/progressive

Compression steps

Res. 0

5 MSB <sub>Z</sub>				
2	5	8	12	20
6	4	9	22	20
1	3	25	20	24
4	11	21	23	20
8	5	2	3	22
10	15	8	11	4
15	12	7	8	4
13	17	15	11	6
5	9	2	5	22
13	17	15	11	6
9	25	5	9	25

4 MSB <sub>Z</sub>				
0	4	6	12	25
4	3	8	22	17
0	3	24	21	22
3	11	21	24	19
6	4	2	3	19
8	14	9	12	3
14	13	7	10	4
13	18	16	13	6
10	26	6	10	26

3 MSB <sub>Z</sub>				
0	0	2	10	12
0	0	6	22	16
0	1	23	21	20
1	11	21	23	20
0	0	0	3	17
3	11	7	11	3
10	12	5	9	3
12	18	17	13	5
7	10	26	10	26

2 MSB <sub>Z</sub>				
0	0	0	12	2
0	4	4	24	6
0	2	22	22	16
2	10	22	22	16
0	0	0	6	8
0	12	6	14	0
6	14	4	8	0
14	22	16	12	2
6	8	8	8	24

Res. -1

4	13	21	28
5	22	22	16
10	6	12	18
14	10	6	15

Sig	Sig	Sig	Ztr	0	Sig
Sig	Sig	Ztr	0	0	0
Ztr	Ztr	Sig	Sig		0 Sig
Ztr	0	0	Ztr		0 Sig

		0	Sig	0	0
		0	0	0	0
		Sig	0	0	0
		0	0	0	0

Res. -2

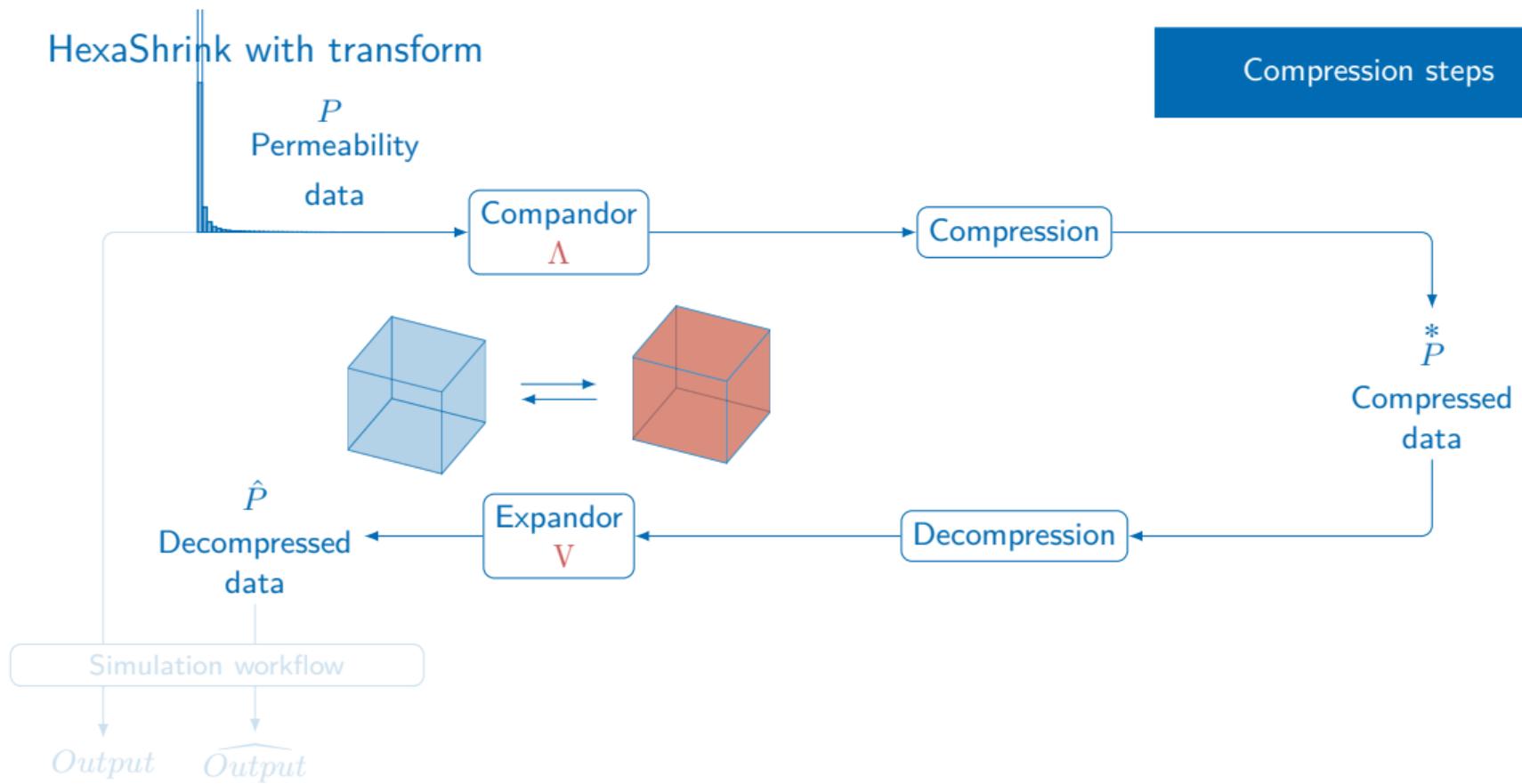
11	22
10	13

Progressive coding with Zerotree:

- EZW (Shapiro, J., 1993)
- SPIHT (Said, A., Pearlman, W., 1996)

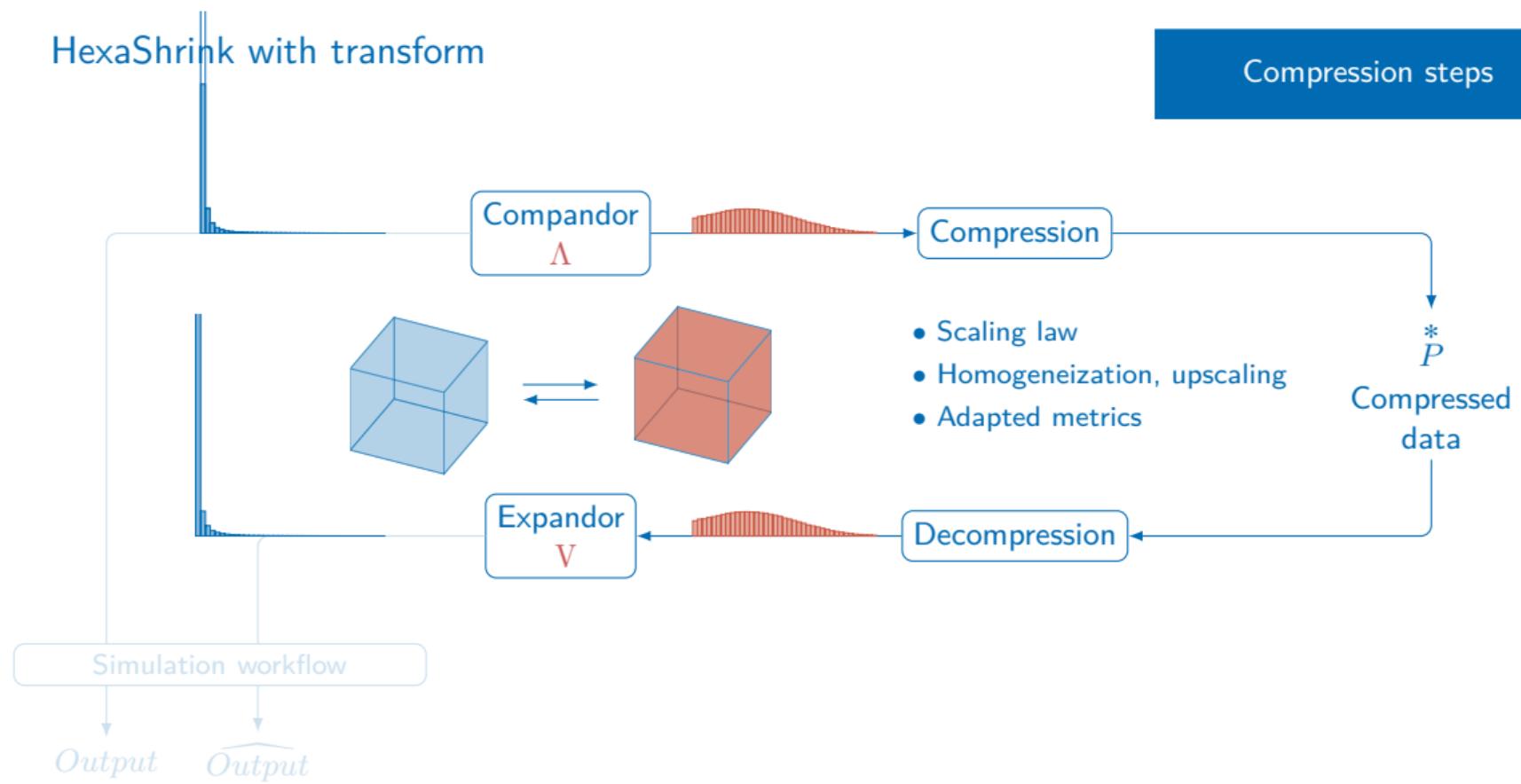
## HexaShrink with transform

Compression steps



## HexaShrink with transform

Compression steps



## Compandor

$$\Lambda_{nbits}(P) = \text{round} \left( \frac{\lambda(P) - \lambda(\min P)}{\lambda(\max P) - \lambda(\min P)} \times (2^{nbits} - 1) \right),$$

with  $\lambda(P) = \log(P + 1)$ ,

## Expandor

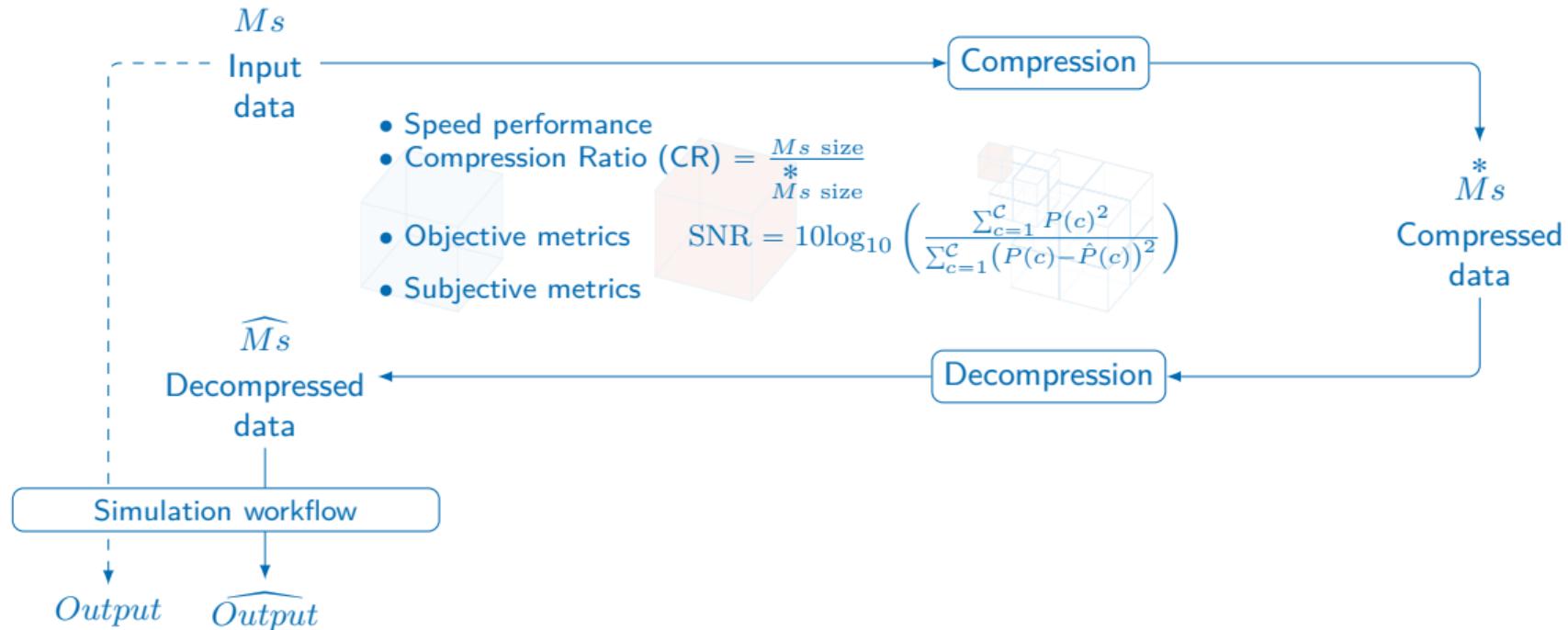
$$\hat{P} = \text{V}_{nbits}(\Lambda_{nbits}(P)) = \text{V} \left( \frac{\Lambda_{nbits}(P)}{(2^{nbits} - 1)} \times (\lambda(\max P) - \lambda(\min P)) + \lambda(\min P) \right),$$

with  $\text{V}(P) = \exp(P) - 1$ .

*Output**Output*

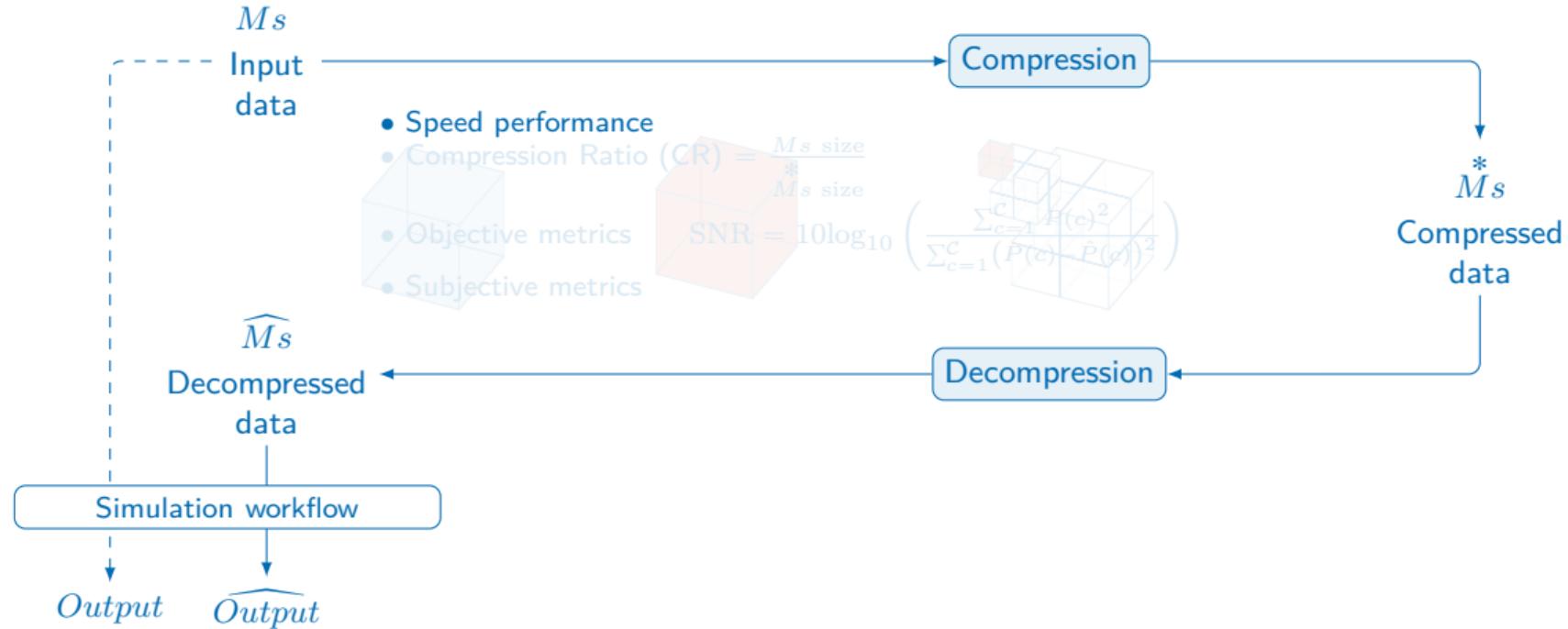
## Evaluation criteria

## Compression steps



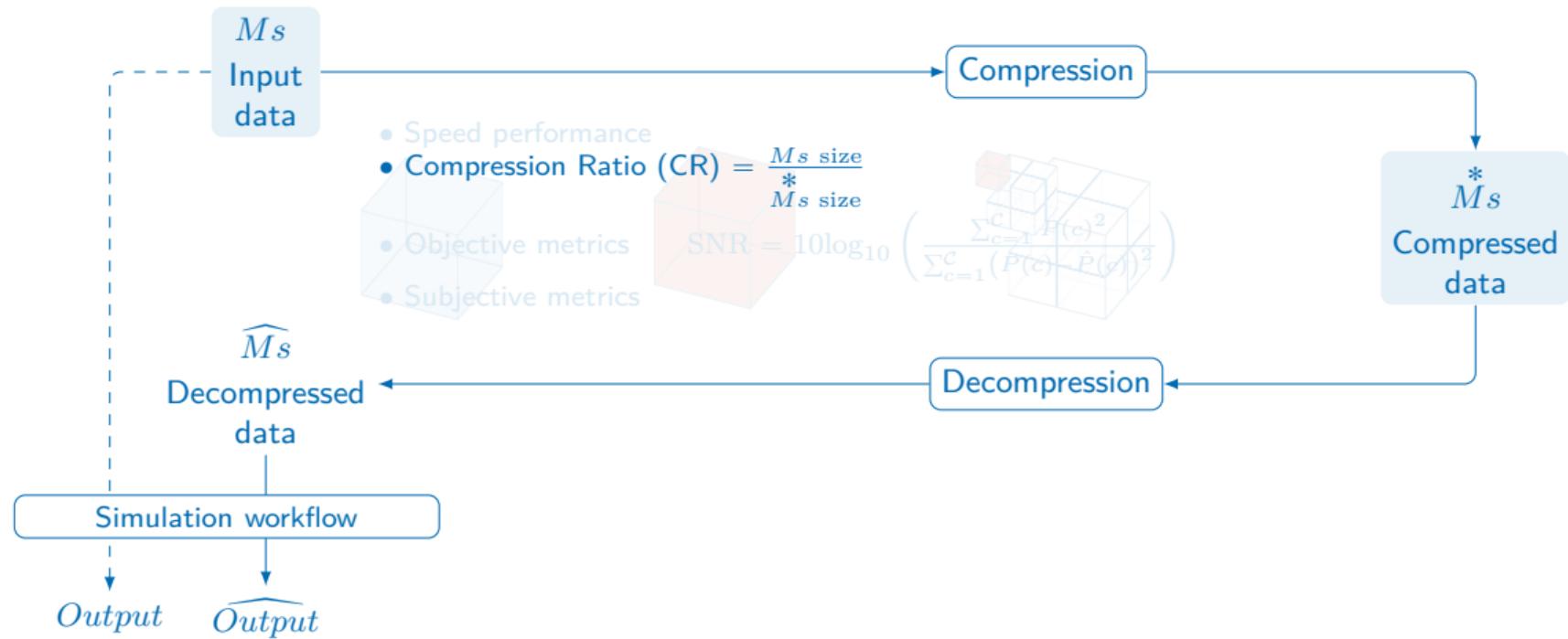
## Evaluation criteria

## Compression steps



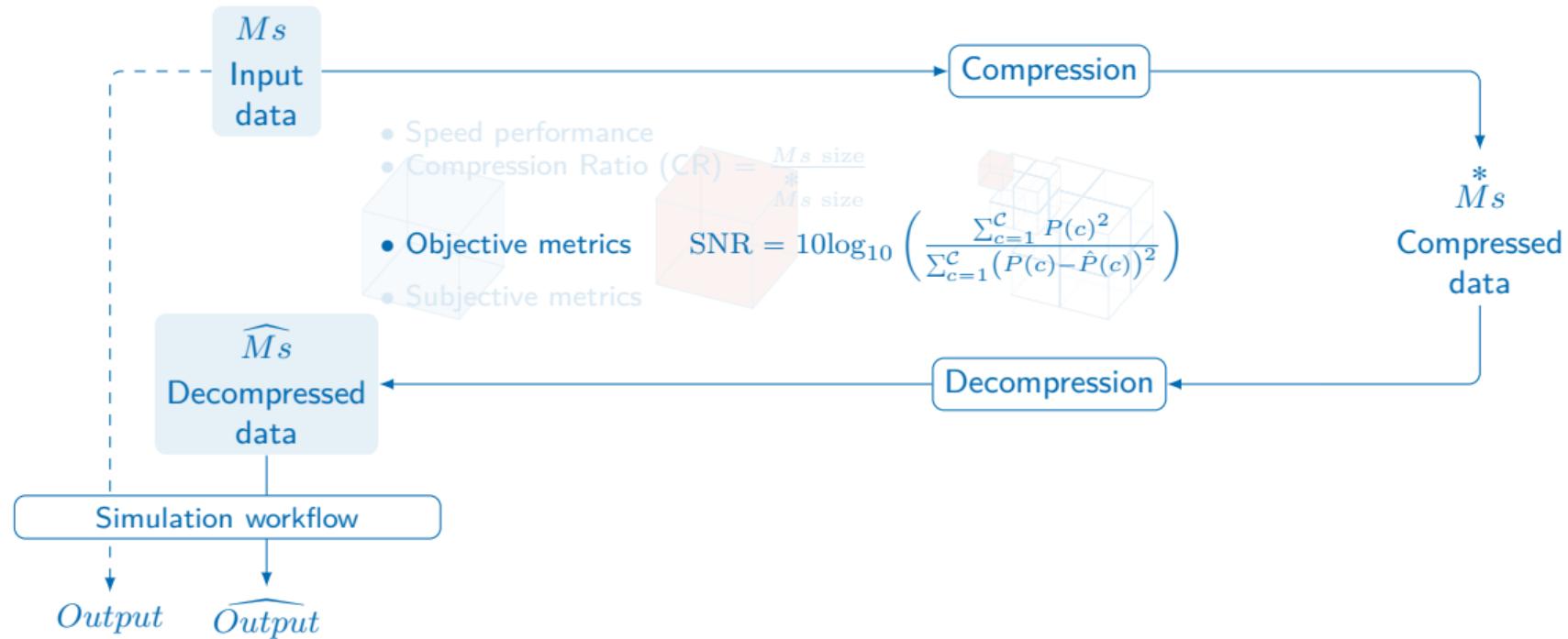
## Evaluation criteria

## Compression steps



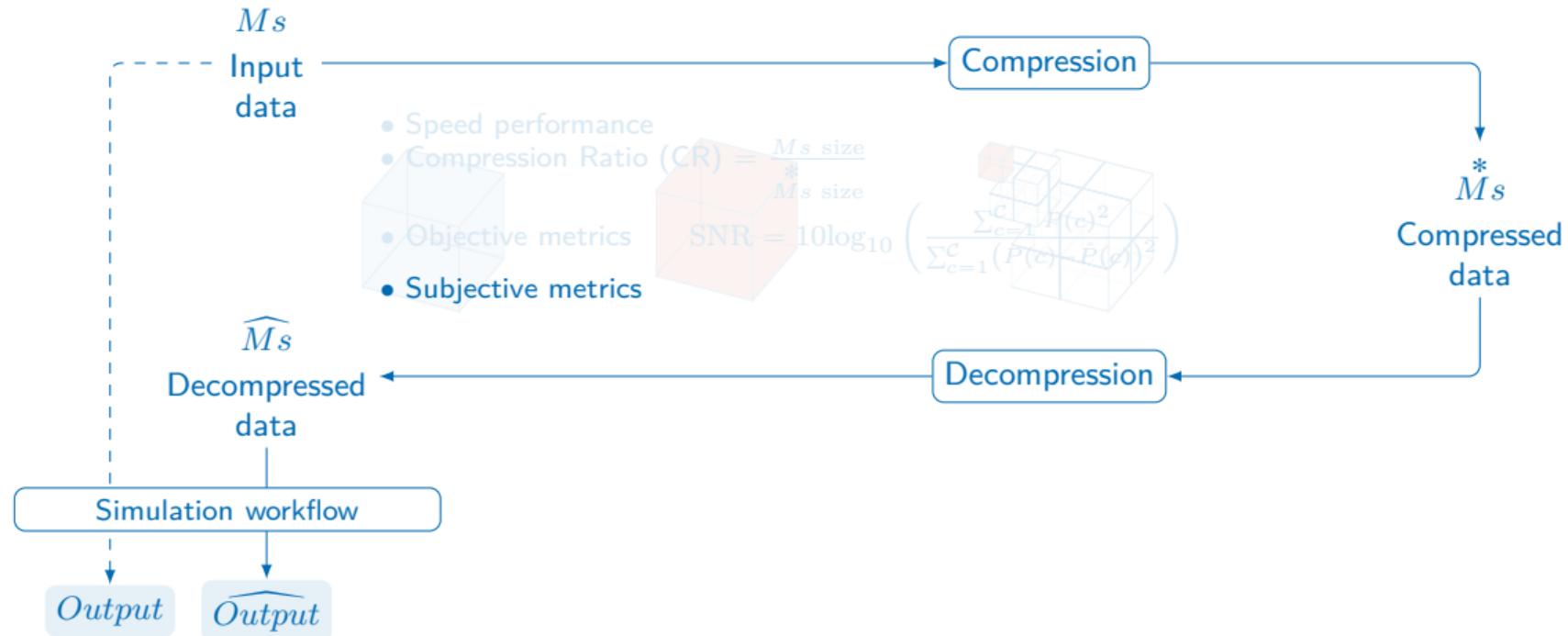
## Evaluation criteria

## Compression steps



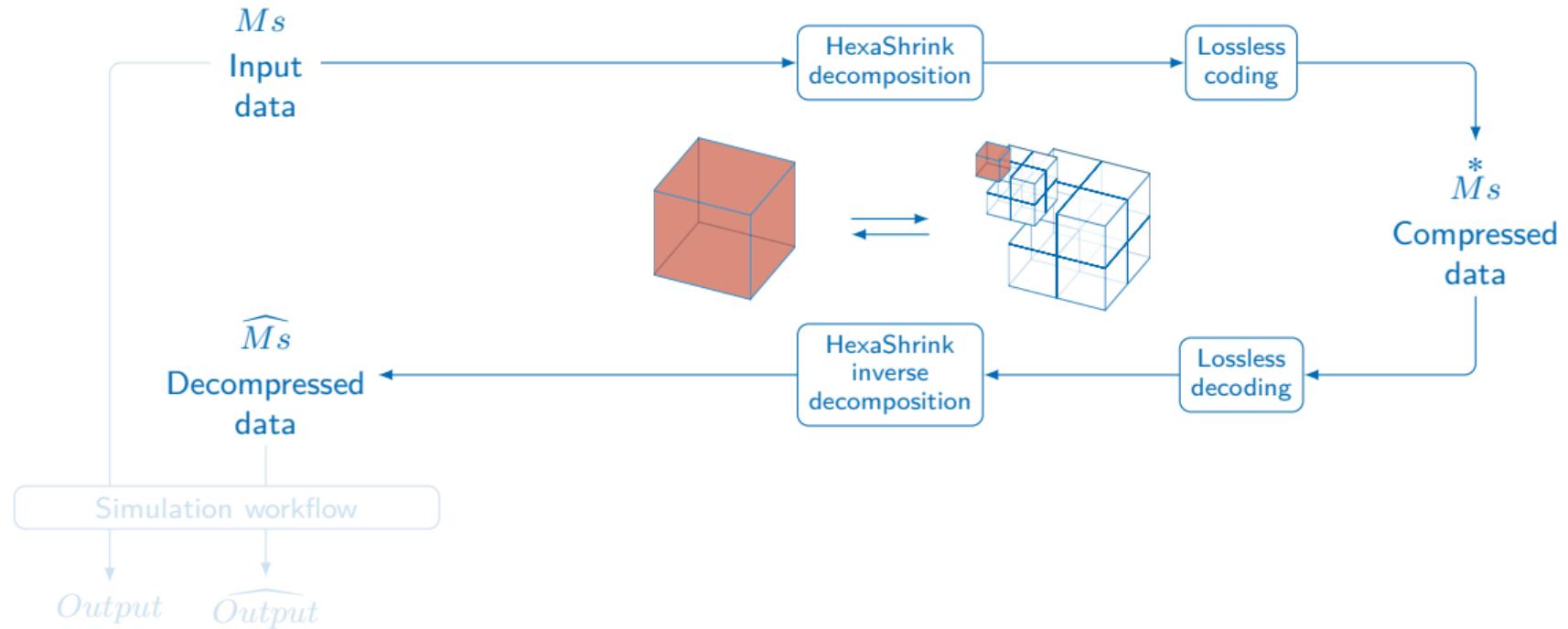
## Evaluation criteria

## Compression steps



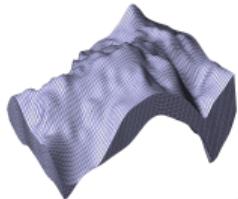
# HexaShrink lossless

Compression (objective)



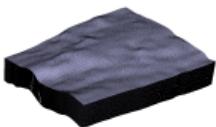
## Benchmark

### Compression (objective)



#1

**faults**  
93,600.00 cells  
4.62 MB



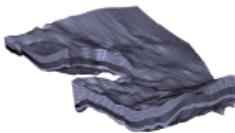
#2

**faults**  
1,000,000.00 cells  
42.46 MB



#3

**faults**  
36,816.00 cells  
1.46 MB



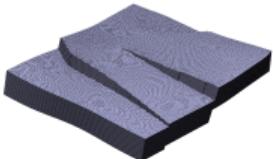
#4

**faults**  
210,000.00 cells  
7.88 MB



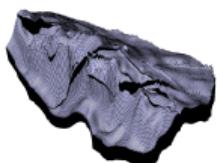
#5

**faults**  
450,576.00 cells  
22.73 MB



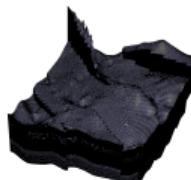
#6

**faults**  
524,288.00 cells  
64.27 MB



#7

**faults**  
6,577,325.00 cells  
274.57 MB



#8

**faults**  
13,947,600.00 cells  
580.94 MB

## HexaShrink lossless performance

Compression (objective)

On mesh#6



Features		Compression ratio			
Faults	File Size	Level	gzip (1992)	bzip2 (1996)	LZMA (1998)
Yes	64.27 MB	none	1.88	2.25	3.04
		1.00	2.70	3.17	3.71
		2.00–6.00	2.84–2.86	3.39–3.42	3.90–3.93

### Observations:

- Lossless coders provide good performance
- HexaShrink improves further compression rate
- Less significant improvements above two levels

# HexaShrink lossless performance

Compression (objective)

On mesh#6



Features		Compression ratio			
Faults	File Size	Level	gzip (1992)	bzip2 (1996)	LZMA (1998)
Yes	64.27 MB	none	1.88	2.25	3.04
		1.00	2.70	3.17	3.71
		2.00–6.00	2.84–2.86	3.39–3.42	3.90–3.93

Observations:

- Lossless coders provide good performance
- HexaShrink improves further compression rate
- Less significant improvements above two levels

## HexaShrink lossless performance

Compression (objective)

On mesh#6



Features		Compression ratio			
Faults	File Size	Level	gzip (1992)	bzip2 (1996)	LZMA (1998)
Yes	64.27 MB	none	1.88	2.25	3.04
		1.00	2.70	3.17	3.71
		2.00–6.00	2.84–2.86	3.39–3.42	3.90–3.93

Observations:

- Lossless coders provide good performance
- HexaShrink improves further compression rate
- Less significant improvements above two levels

## HexaShrink lossless performance

Compression (objective)

On mesh#6



Features		Compression ratio			
Faults	File Size	Level	gzip (1992)	bzip2 (1996)	LZMA (1998)
Yes	64.27 MB	none	1.88	2.25	3.04
		1.00	2.70	3.17	3.71
		2.00–6.00	2.84–2.86	3.39–3.42	3.90–3.93

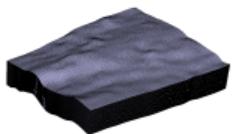
Observations:

- Lossless coders provide good performance
- HexaShrink improves further compression rate
- Less significant improvements above two levels

# HexaShrink lossless performance

Compression (objective)

On all meshes

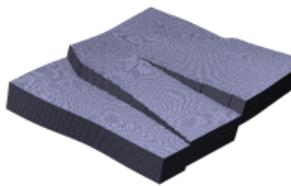
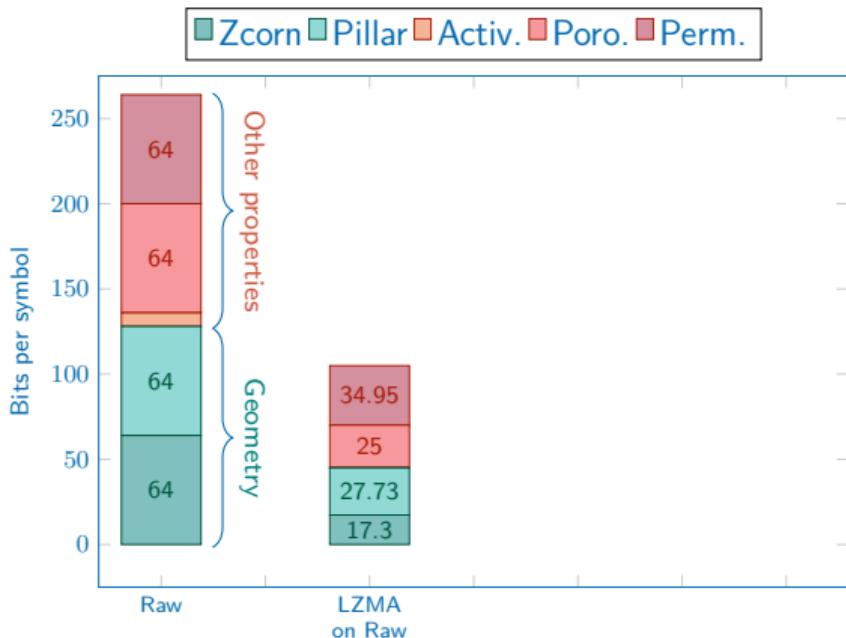


Mesh	Features		Level	Compression ratio		
	Faults	File Size		gzip (1992)	bzip2 (1996)	LZMA (1998)
1	No	4.62 MB	none	3.73	4.98	6.43
			1	5.62	6.07	7.52
			2.00–4.00	<b>5.67</b>	<b>6.12–6.13</b>	<b>7.42–7.44</b>
2	No	42.46 MB	none	3.23	8.41	10.12
			1	6.49	10.82	11.81
			2.00–6.00	<b>7.48–7.58</b>	<b>12.75–13.03</b>	<b>13.35</b>
3	Yes	1.46 MB	none	2.67	2.99	3.63
			1	3.88	4.70	5.24
			2.00–4.00	<b>4.03–4.05</b>	<b>4.92–4.93</b>	<b>5.47–5.48</b>
4	Yes	7.88 MB	none	1.83	1.89	2.21
			1	2.64	3.06	3.48
			2.00–4.00	<b>2.76</b>	<b>3.22–3.23</b>	<b>3.64–3.65</b>
5	Yes	22.73 MB	none	2.46	2.55	3.33
			1	3.14	2.83	3.71
			2.00–4.00	<b>3.25–3.26</b>	<b>2.91–2.92</b>	<b>3.80–3.81</b>
6	Yes	64.27 MB	none	1.88	2.25	3.04
			1	2.70	3.17	3.71
			2.00–6.00	<b>2.84–2.86</b>	<b>3.39–3.42</b>	<b>3.90–3.93</b>
7	Yes	274.57 MB	none	2.31	2.25	3.04
			1	3.31	3.53	4.44
			2.00–6.00	<b>4.14–4.24</b>	<b>4.48–4.68</b>	<b>5.54–5.73</b>
8	Yes	580.94 MB	none	3.20	5.98	12.52
			1	5.42	7.07	8.90
			2.00–7.00	<b>5.80–6.72</b>	<b>7.63–10.12</b>	<b>9.05–10.23</b>

# HexaShrink lossless performance

In-depth analysis for *LZMA*

Compression (objective)



LZMA efficiency on all components

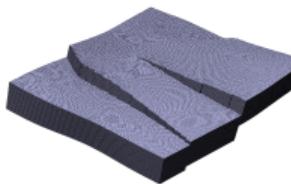
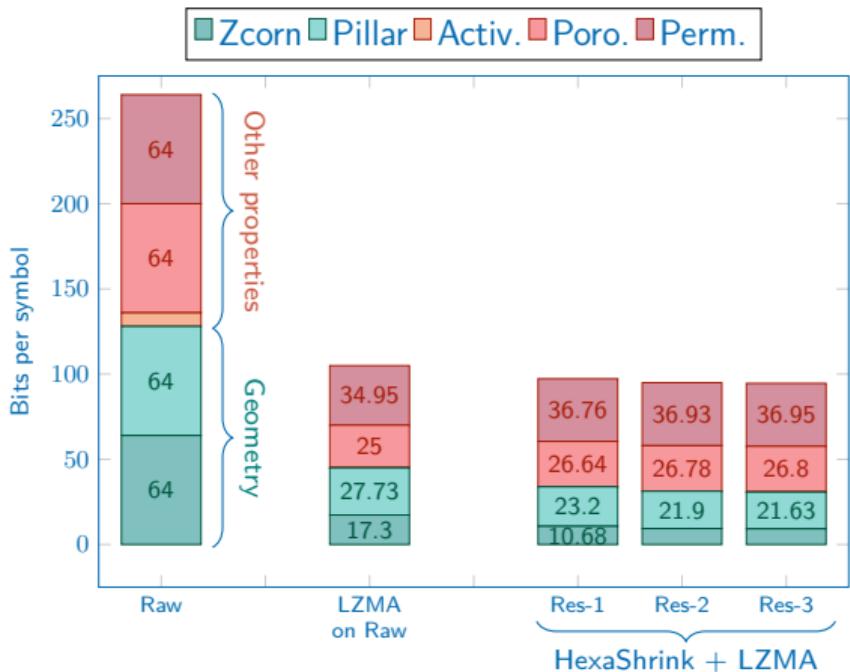
With Hexashrink decomposition:

- Geometry
- Efficient
- Other properties

# HexaShrink lossless performance

In-depth analysis for LZMA

Compression (objective)



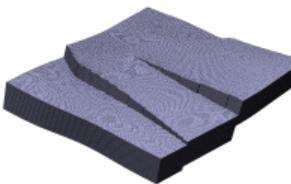
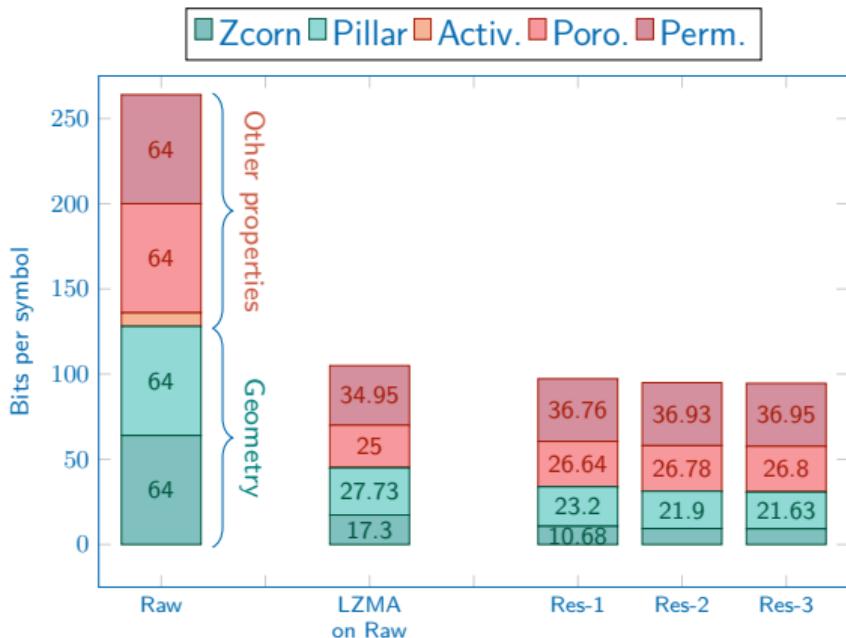
LZMA efficiency on all components  
With Hexashrink decomposition:

- Geometry Efficient
- Other properties Limited

# HexaShrink lossless performance

In-depth analysis for *LZMA*

Compression (objective)



LZMA efficiency on all components

With Hexashrink decomposition:

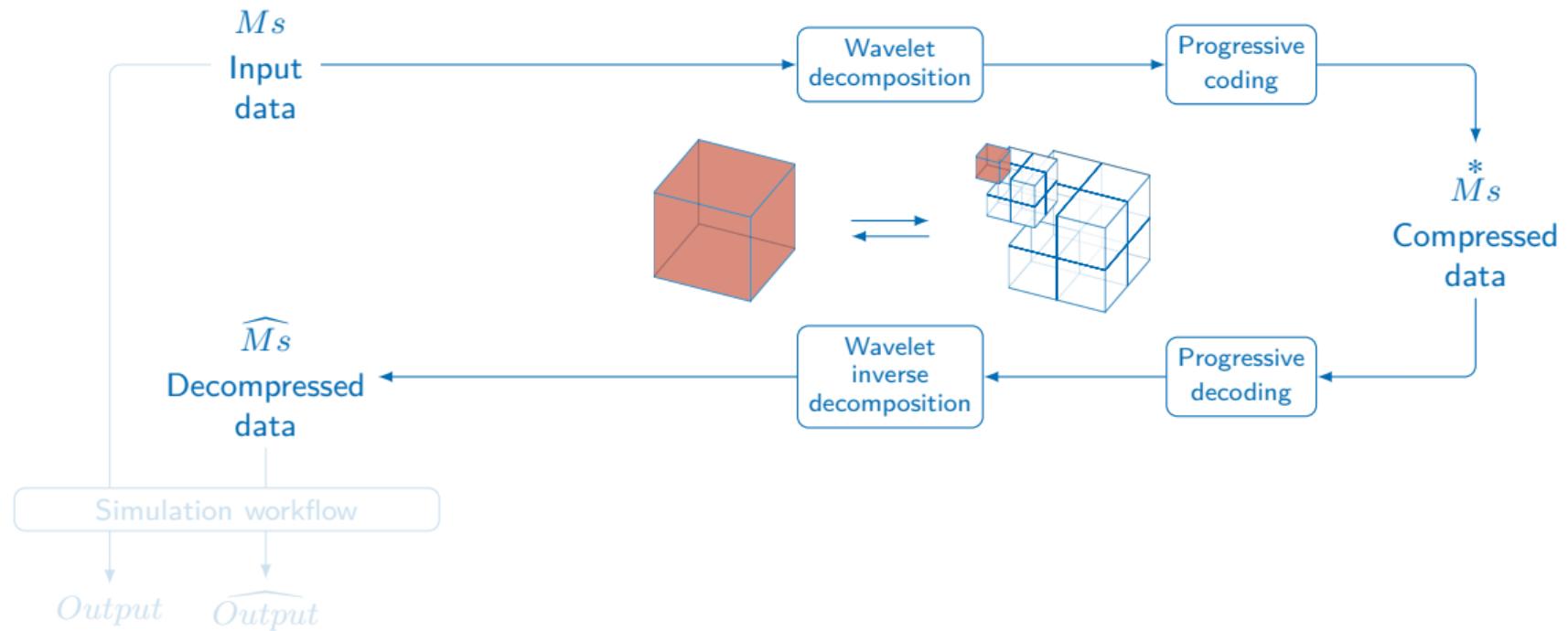
- Geometry

- Other properties

Let us explore refinable precision!

# HexaShrink lossy/progressive performance

Compression (subjective)

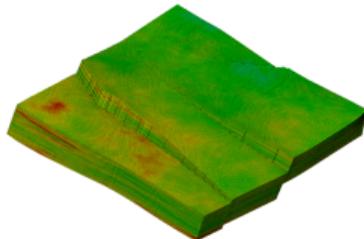
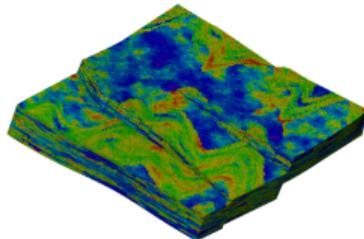
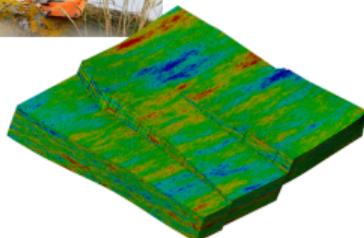
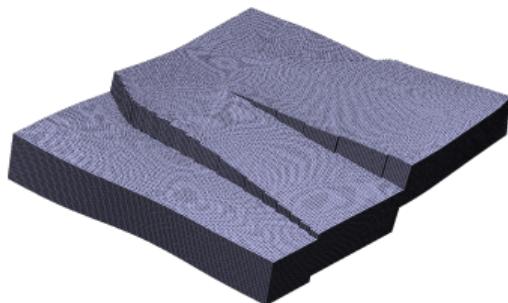
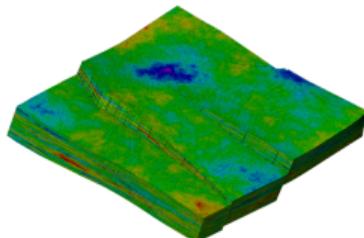


## HexaShrink lossy/progressive performance on Lundi

Four property environments (SPE10, Christie, M. 2001)



Compression (subjective)

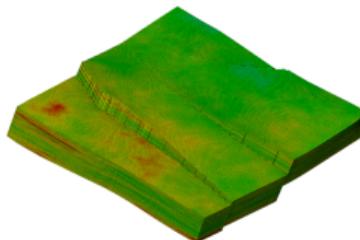
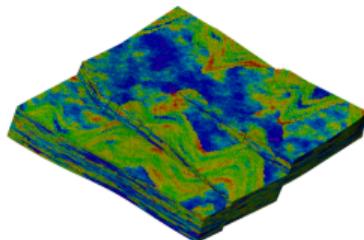
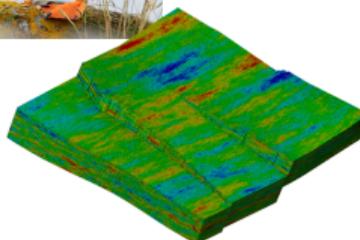
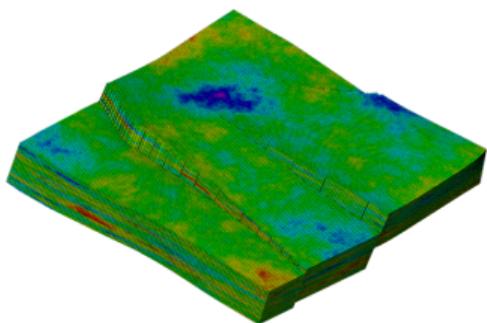
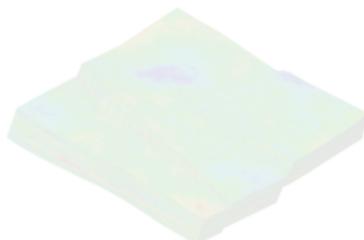


## HexaShrink lossy/progressive performance on Lundi

Four property environments (SPE10, Christie, M. 2001)

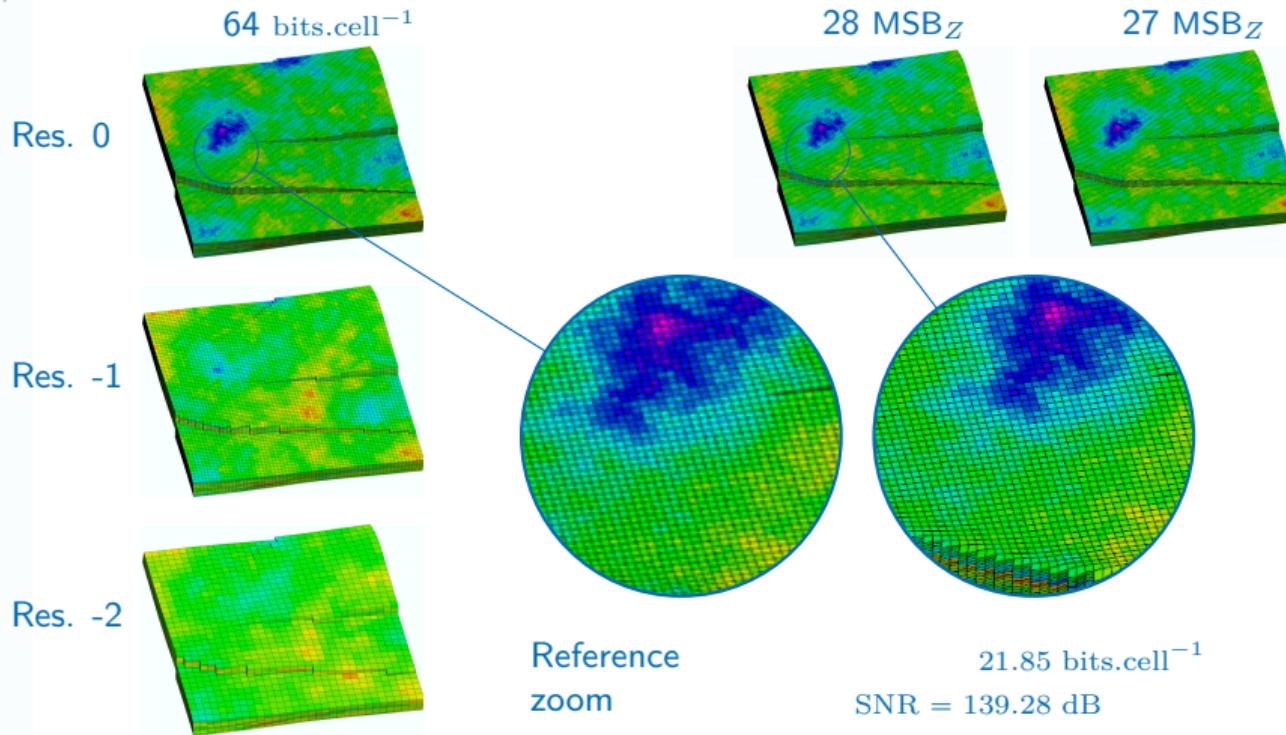


Compression (subjective)



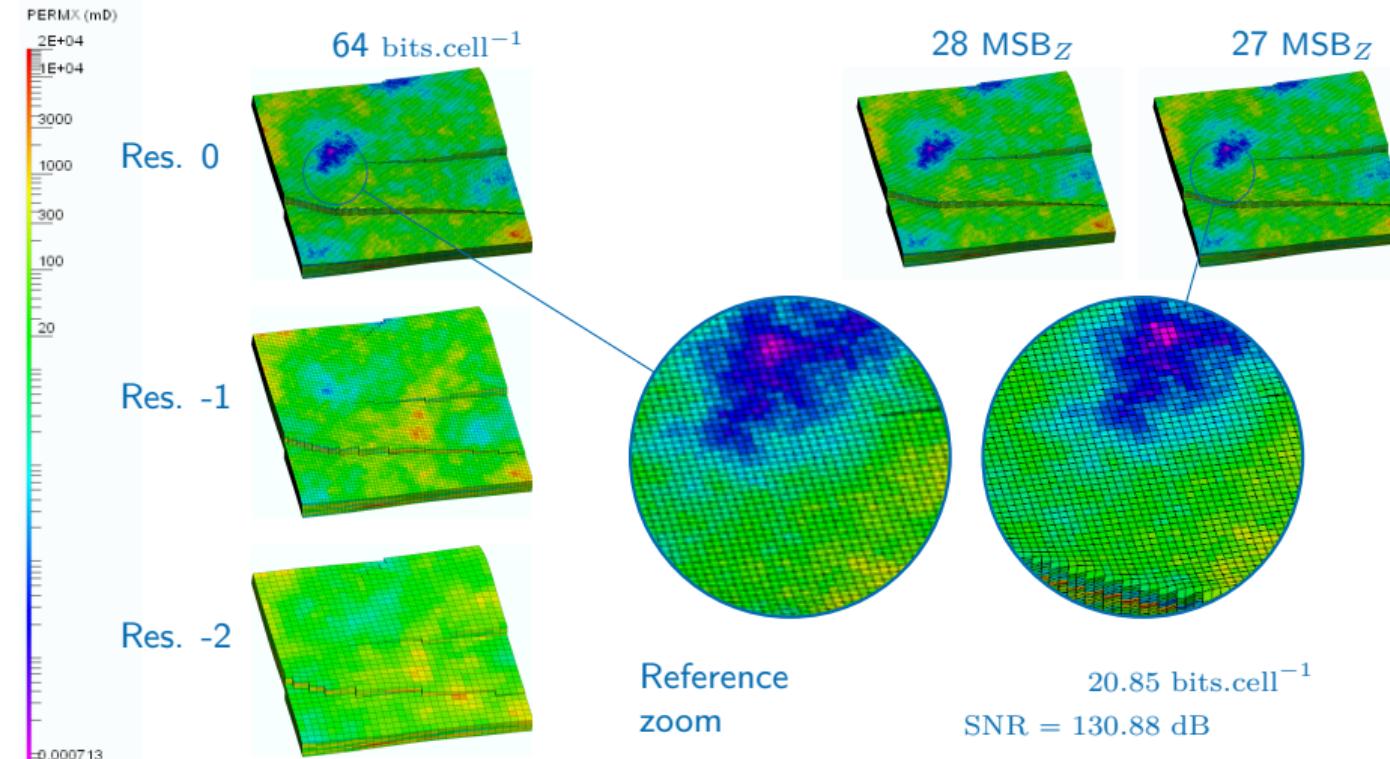
# HexaShrink lossy/progressive: visual results

Compression (subjective)



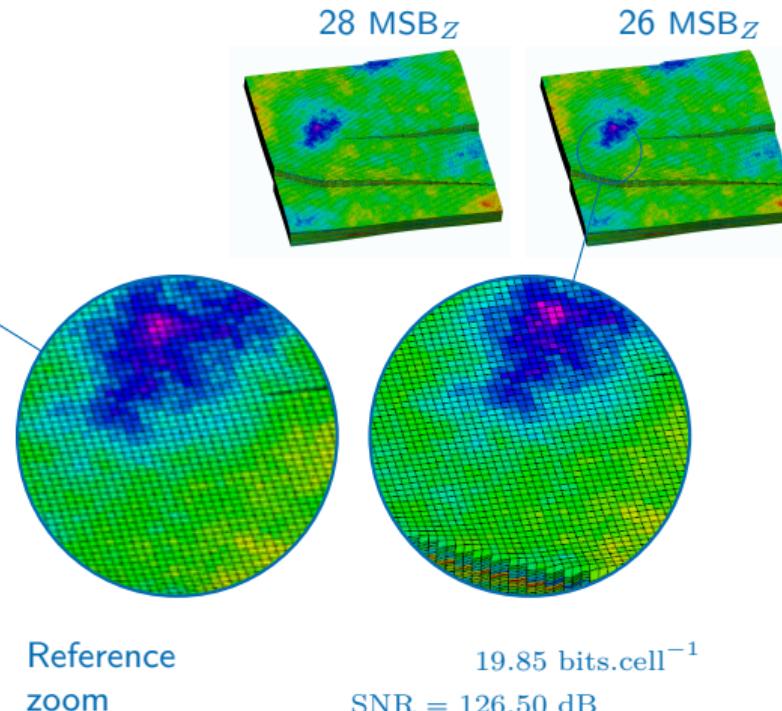
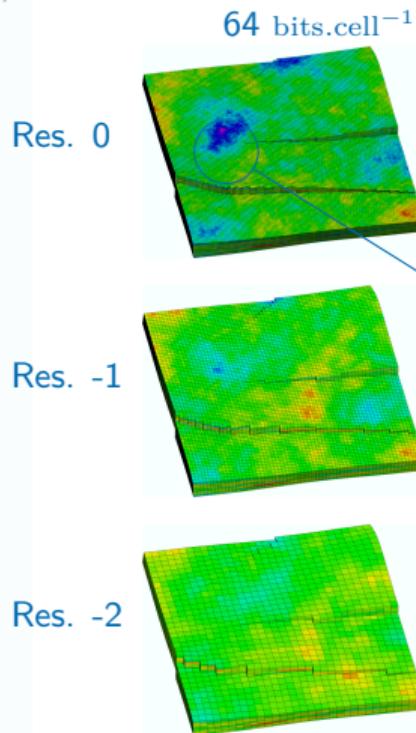
# HexaShrink lossy/progressive: visual results

Compression (subjective)



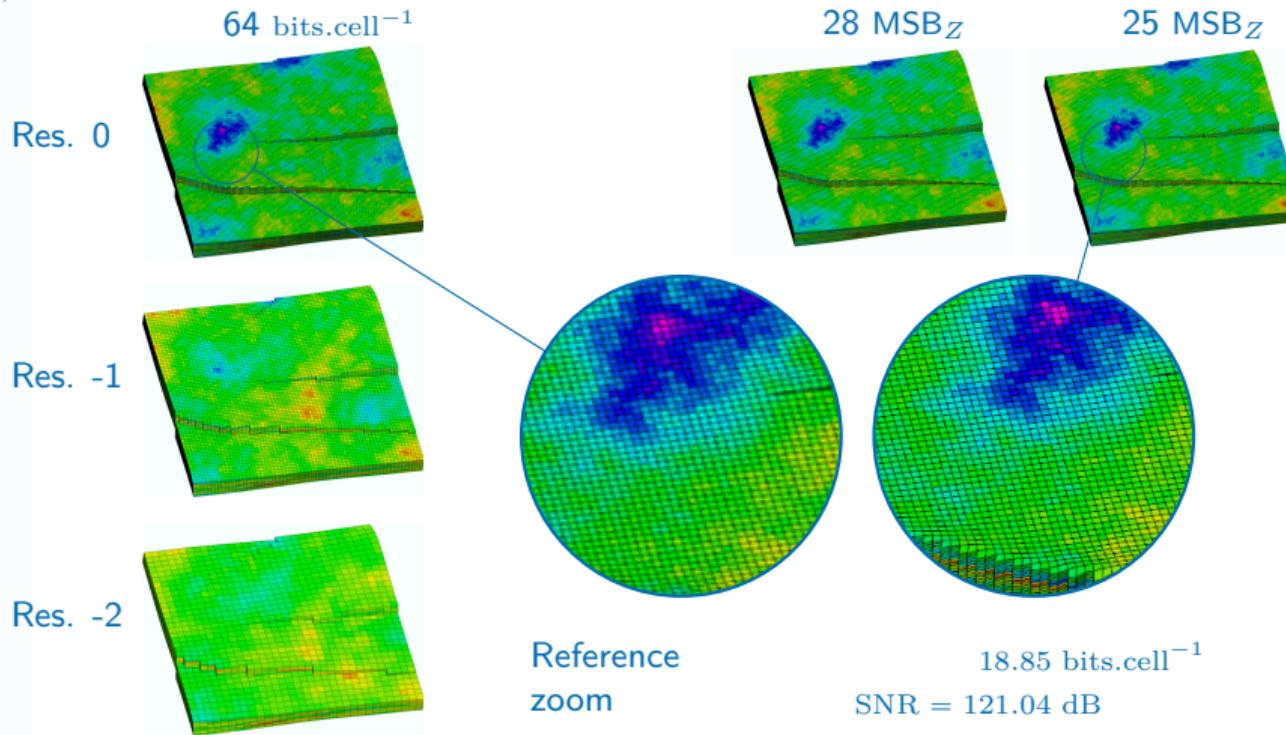
# HexaShrink lossy/progressive: visual results

Compression (subjective)



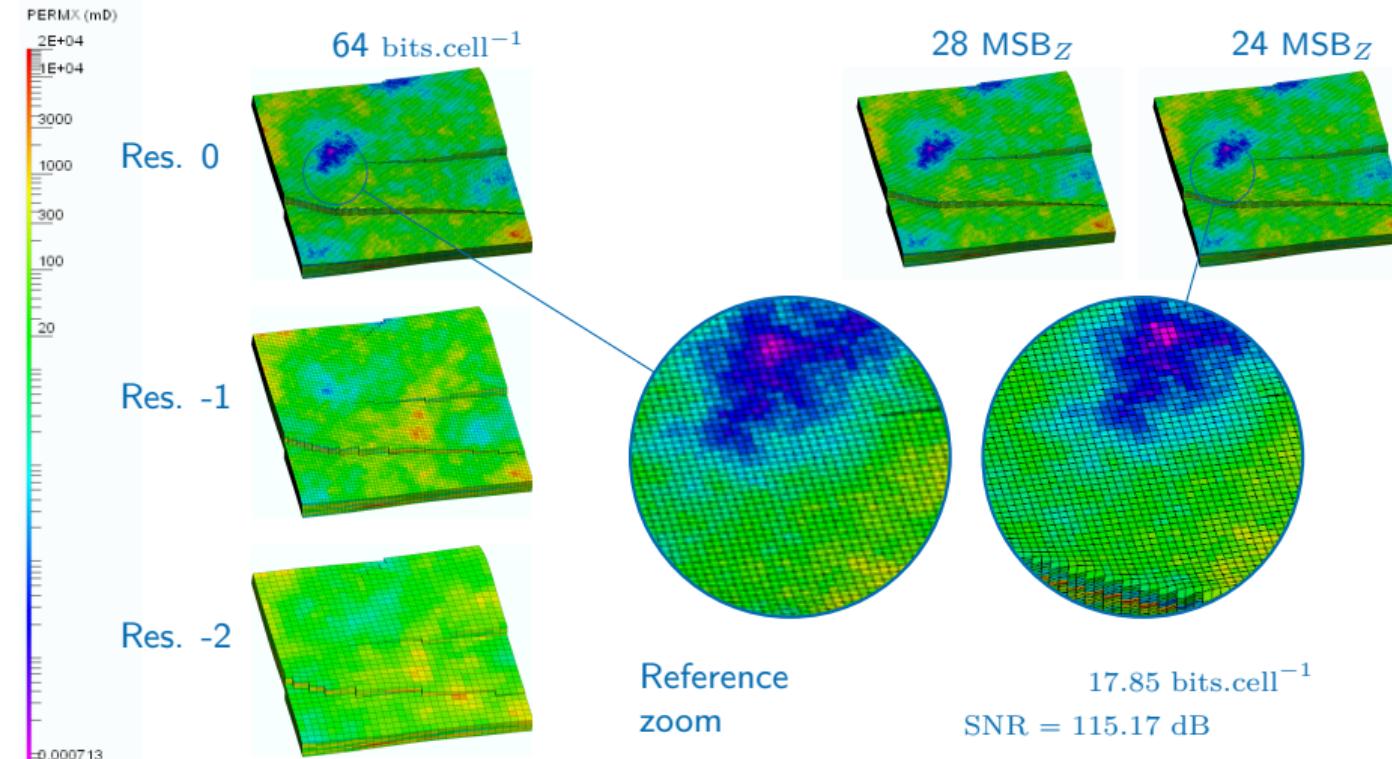
# HexaShrink lossy/progressive: visual results

Compression (subjective)



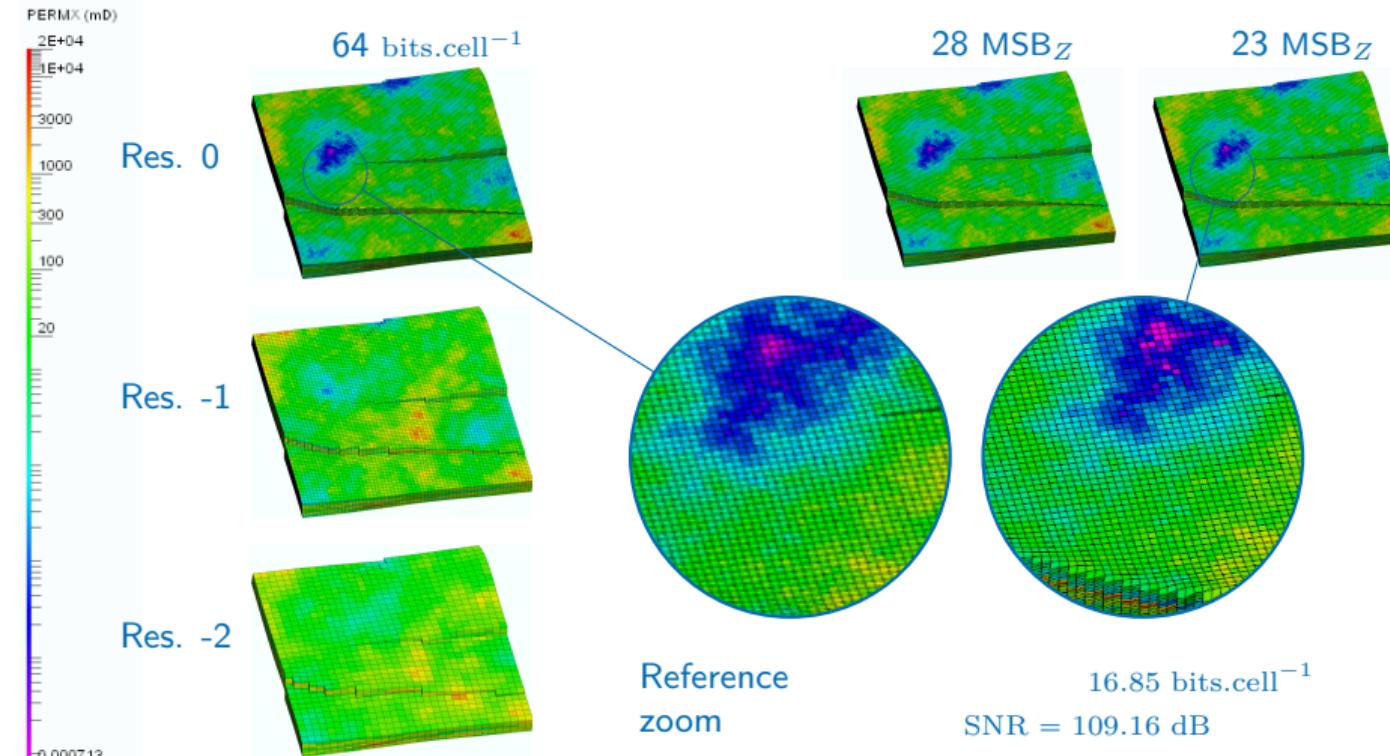
# HexaShrink lossy/progressive: visual results

Compression (subjective)



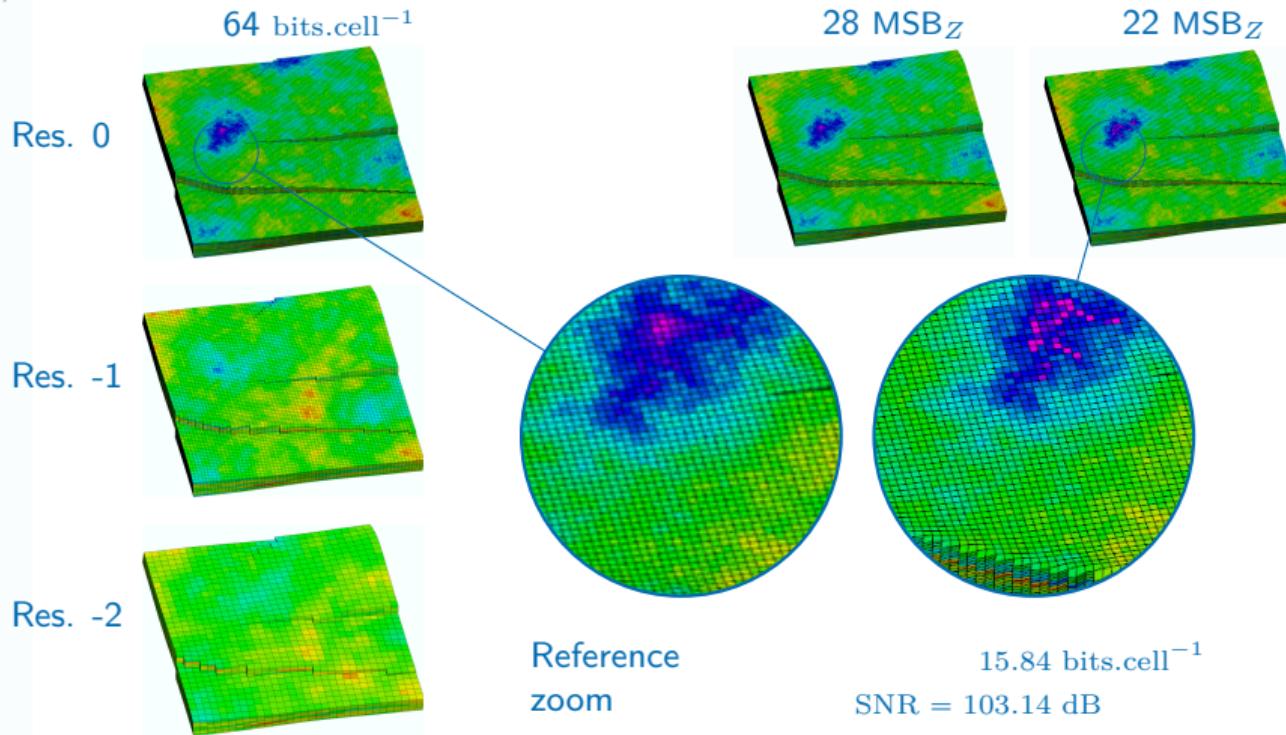
# HexaShrink lossy/progressive: visual results

Compression (subjective)



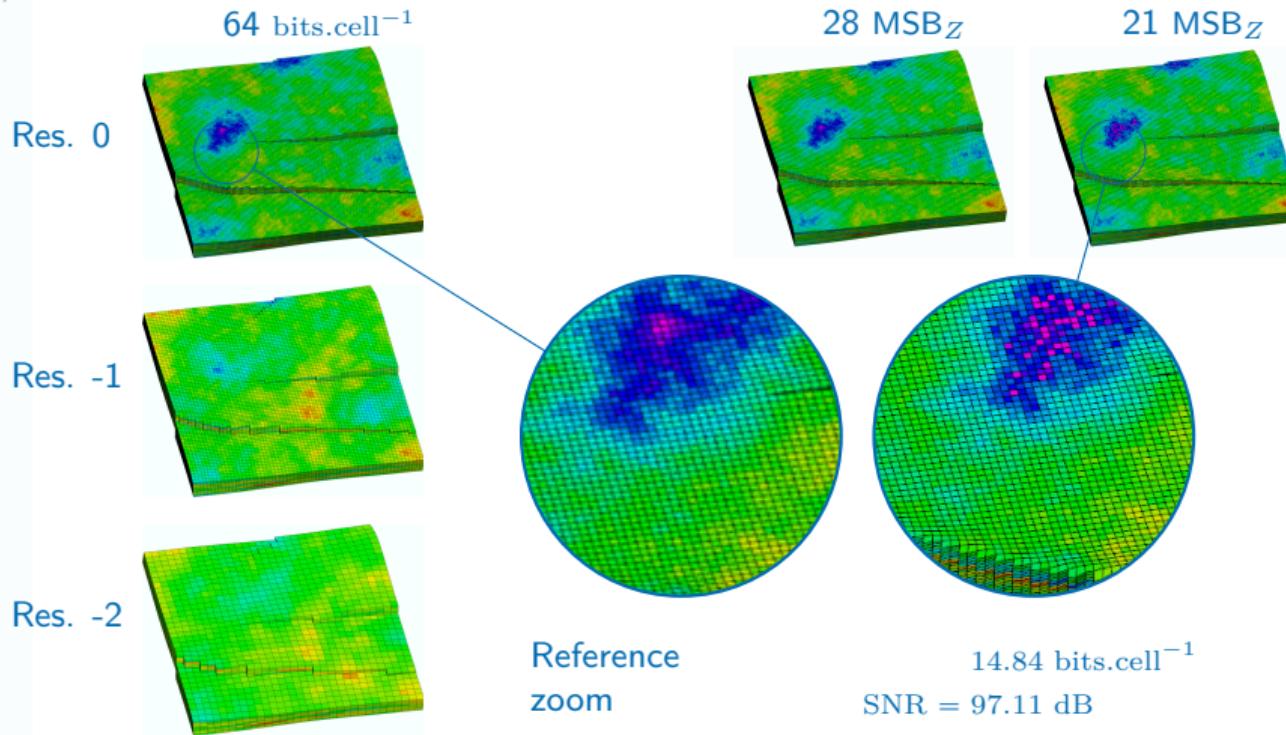
# HexaShrink lossy/progressive: visual results

Compression (subjective)



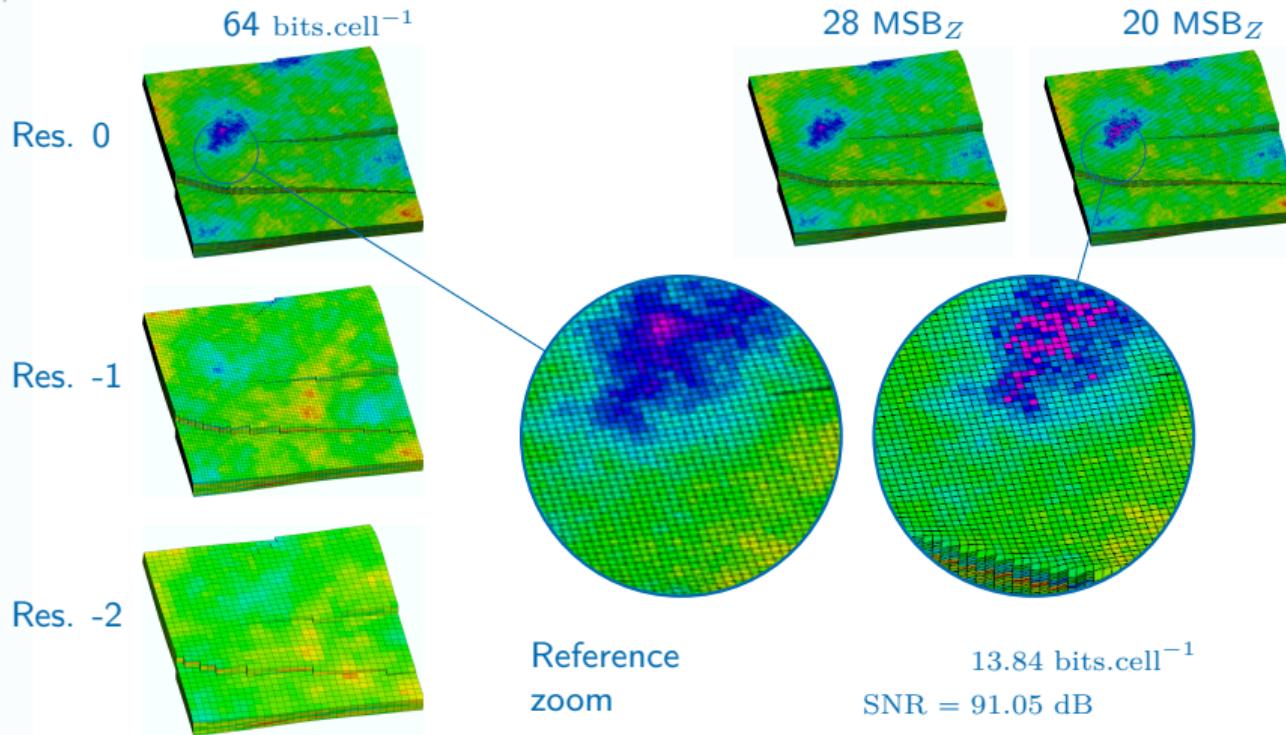
# HexaShrink lossy/progressive: visual results

Compression (subjective)



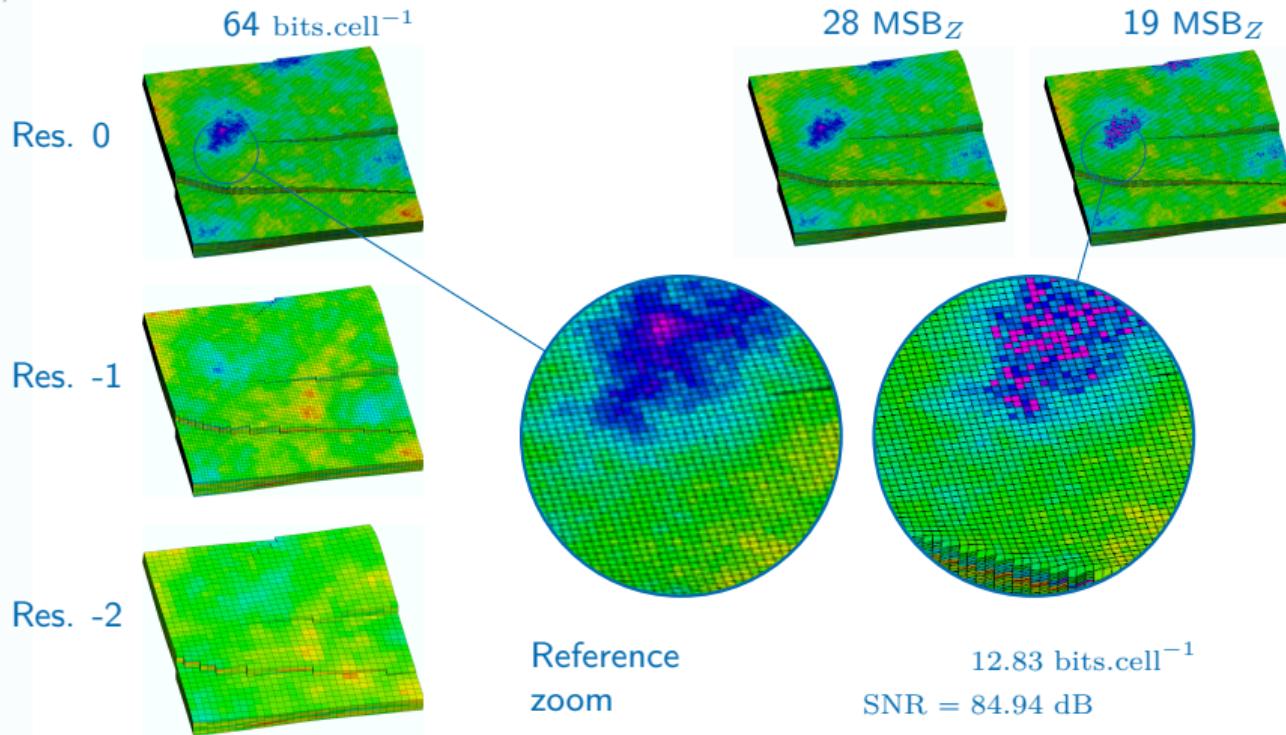
# HexaShrink lossy/progressive: visual results

Compression (subjective)



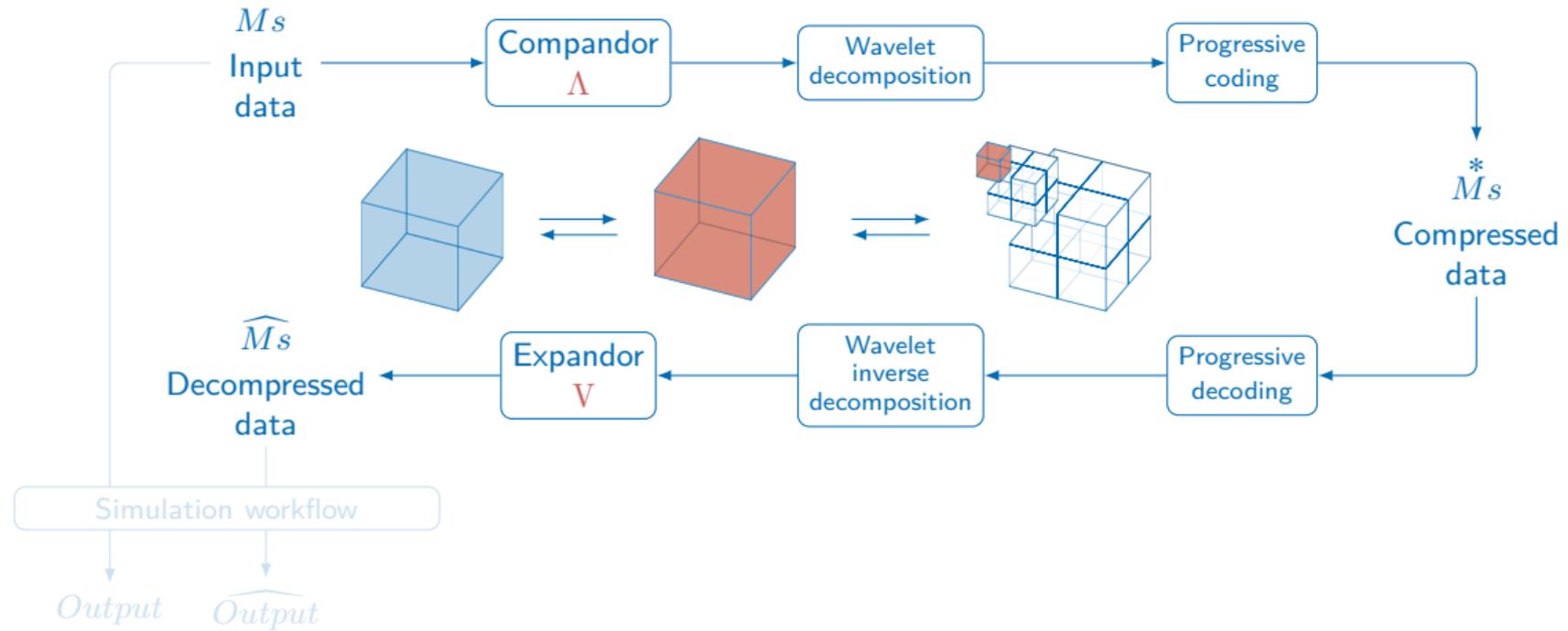
# HexaShrink lossy/progressive: visual results

Compression (subjective)



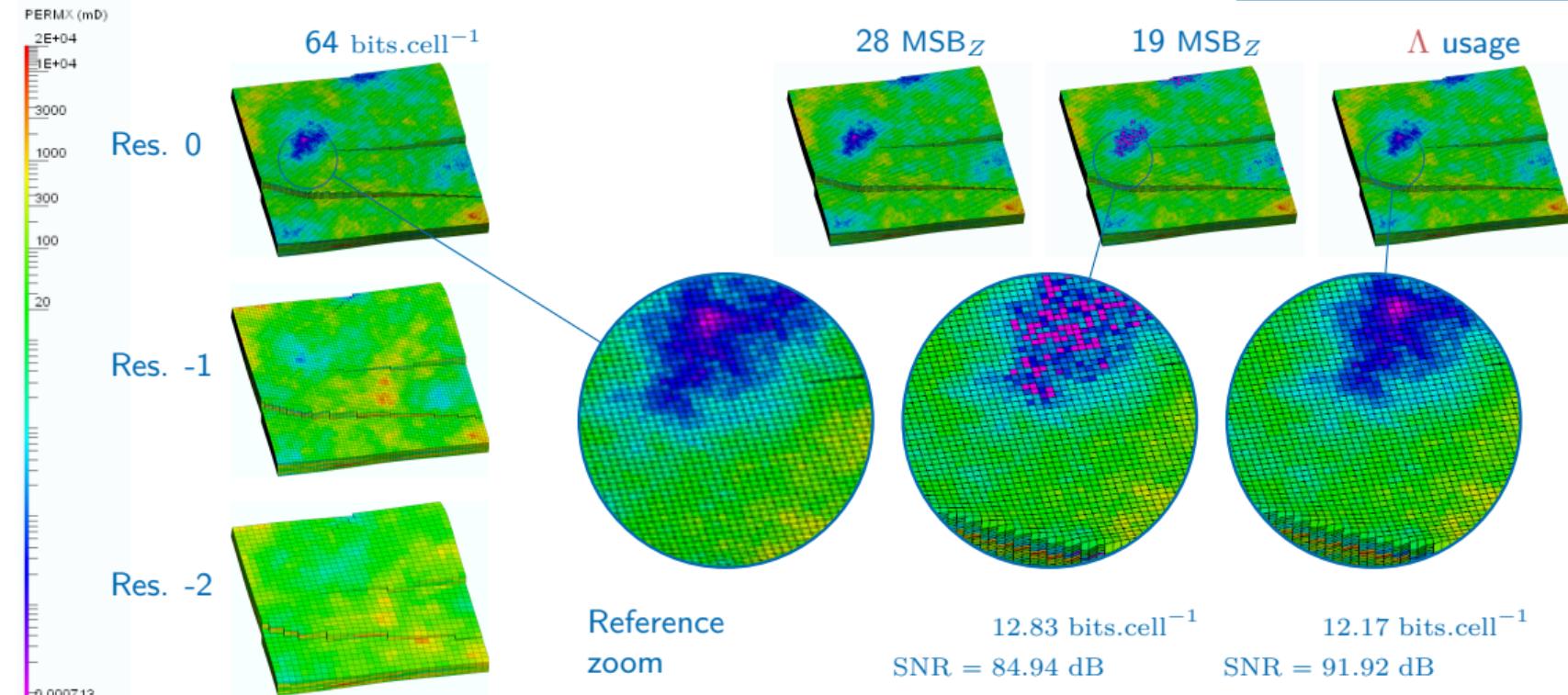
# HexaShrink lossy/progressive with companding: visual results

Compression (subjective)



## 6.0 Compandor effect: visual results

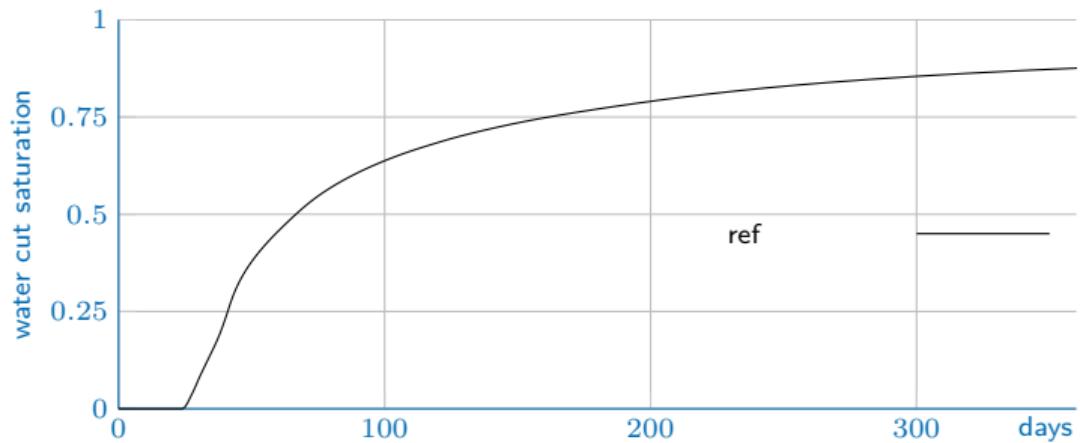
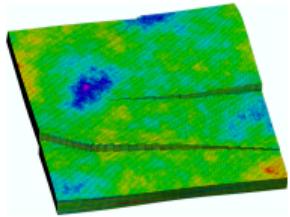
Compression (subjective)



## Simulation: fidelity versus resolution (HexaShrink)

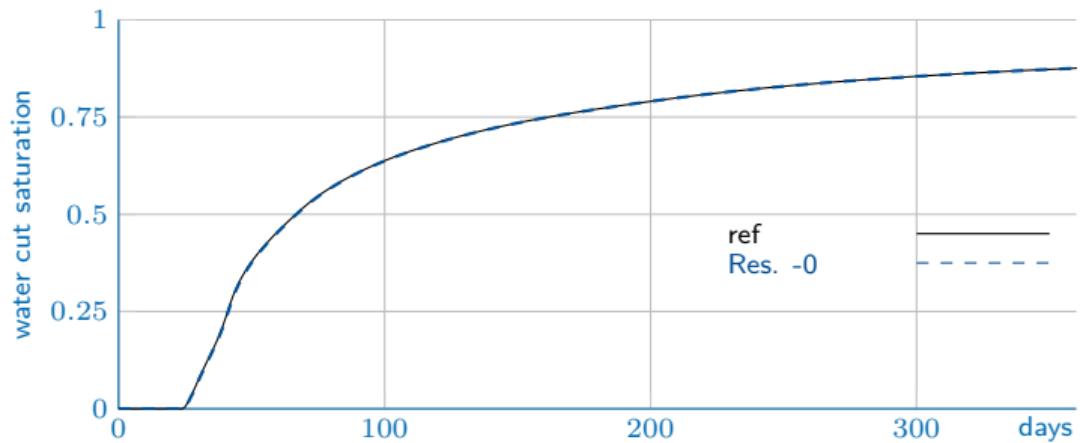
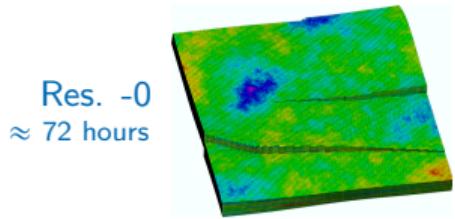
Simulation (subjective)

Raw



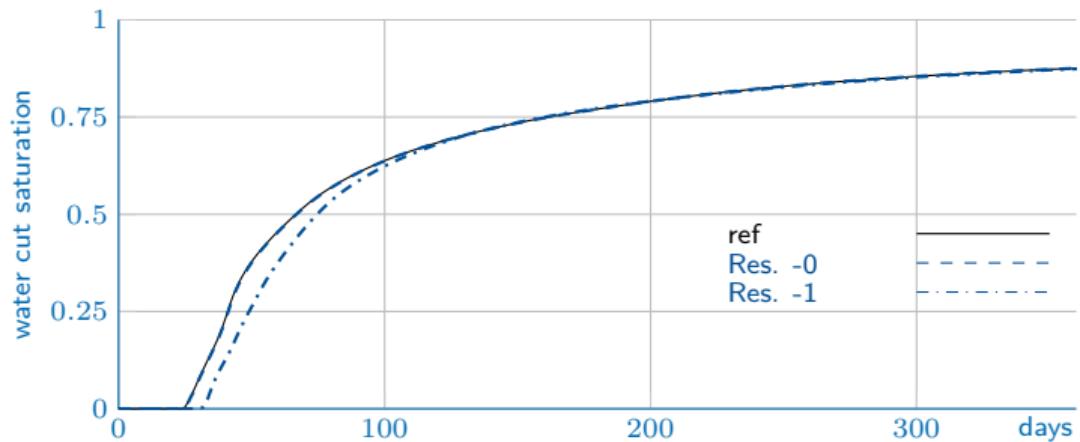
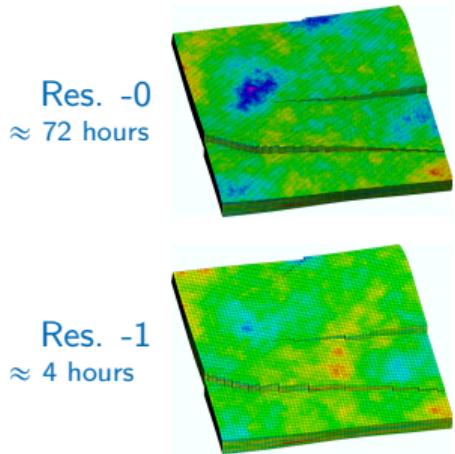
## Simulation: fidelity versus resolution (HexaShrink)

Simulation (subjective)



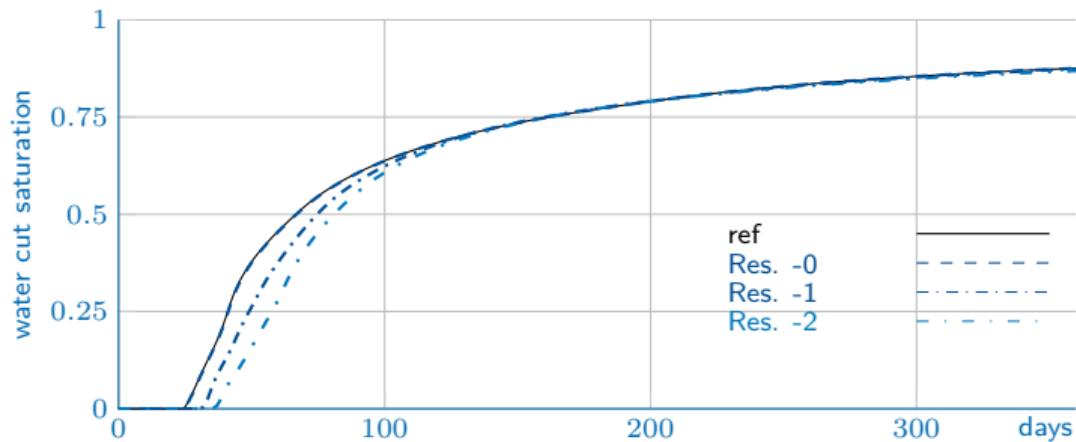
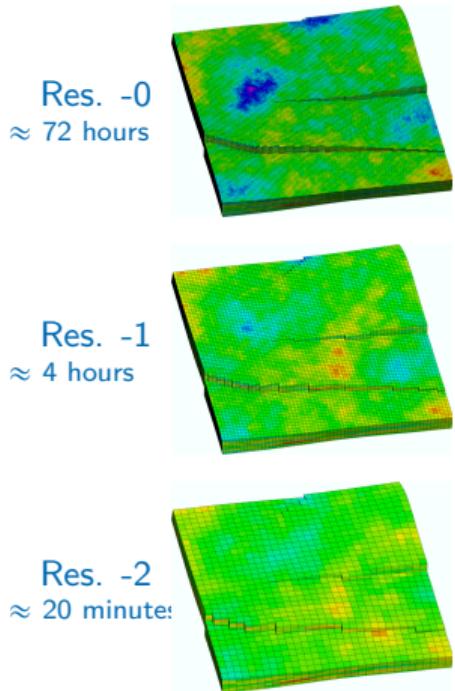
# Simulation: fidelity versus resolution (HexaShrink)

Simulation (subjective)



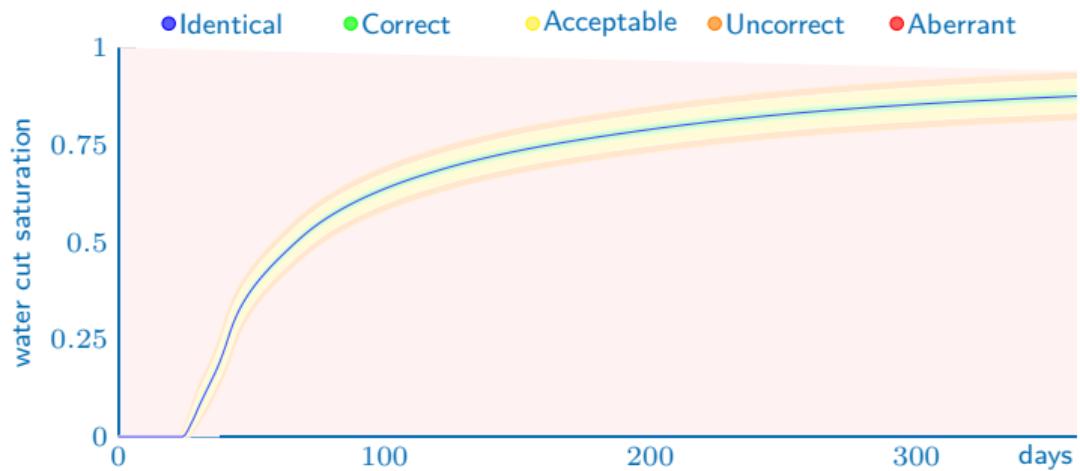
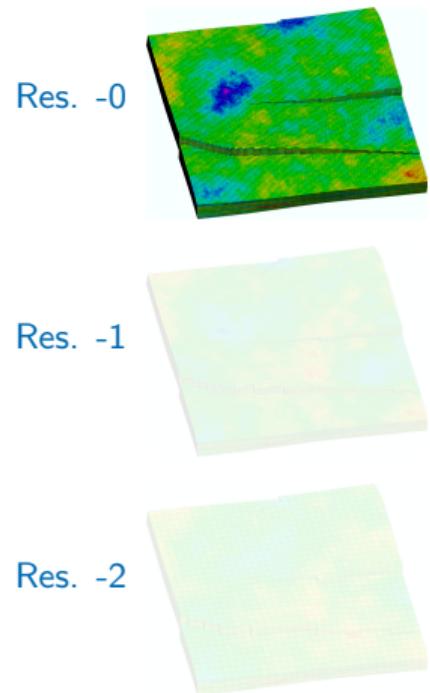
# Simulation: fidelity versus resolution (HexaShrink)

Simulation (subjective)



## Simulation: fidelity versus resolution (HexaShrink)

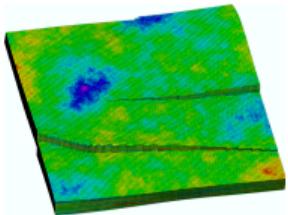
Simulation (subjective)



## Simulation: accuracy versus precision

Simulation (subjective)

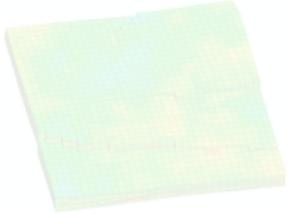
Res. 0



Res. -1

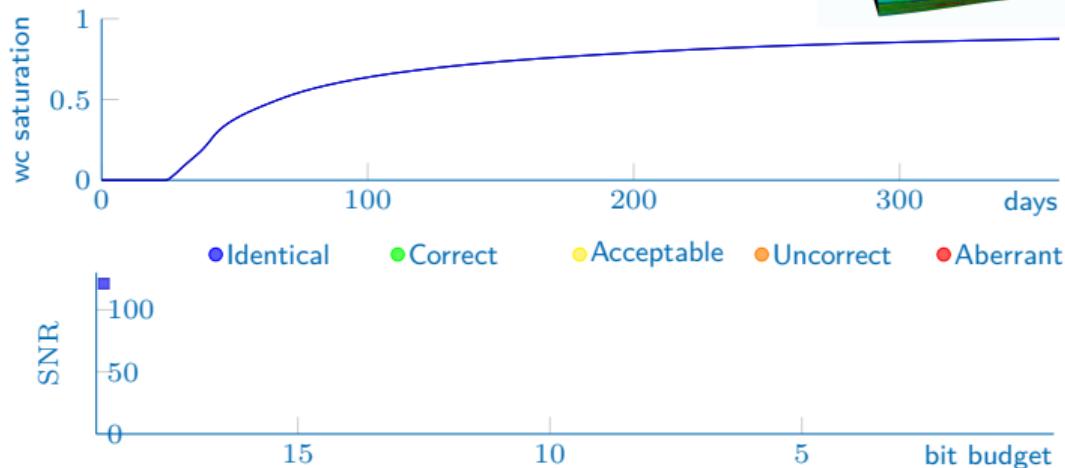
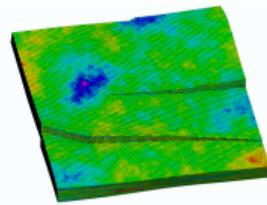


Res. -2



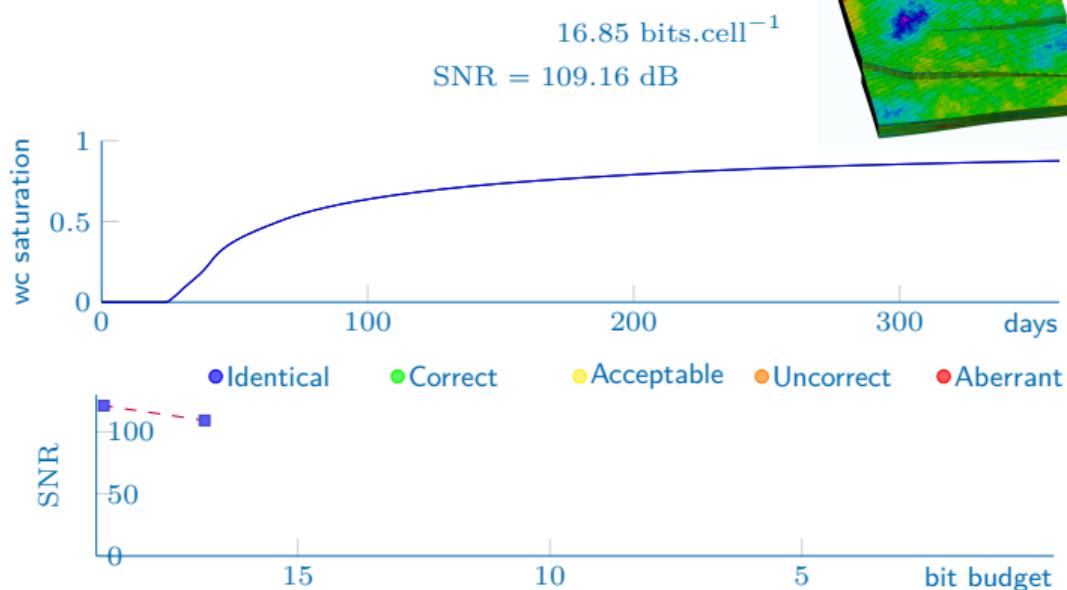
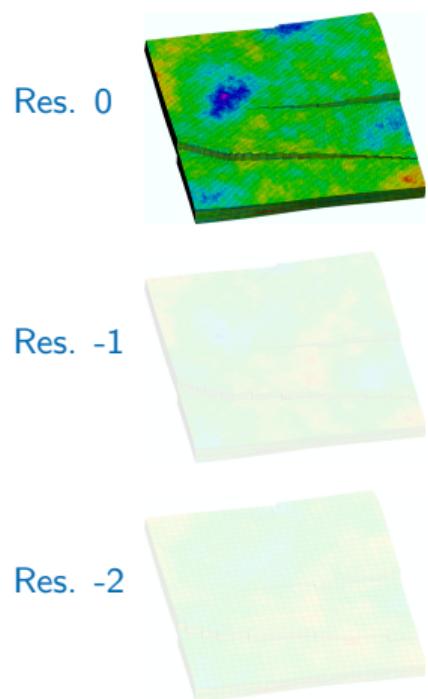
$18.85 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 121.04 \text{ dB}$

$25 \text{ MSB}_Z$



## Simulation: accuracy versus precision

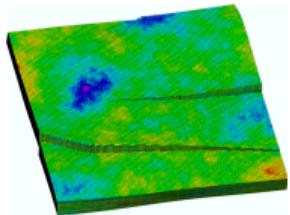
Simulation (subjective)



## Simulation: accuracy versus precision

Simulation (subjective)

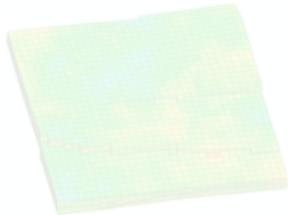
Res. 0



Res. -1

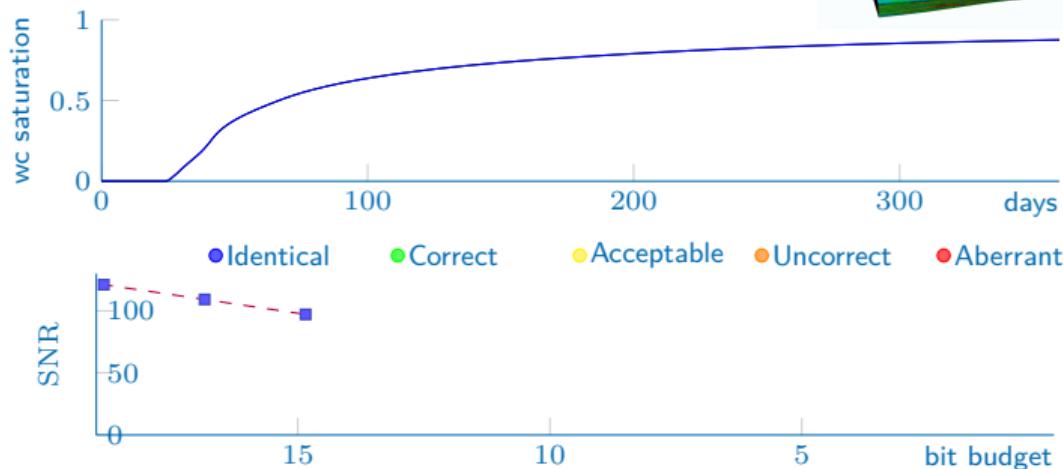
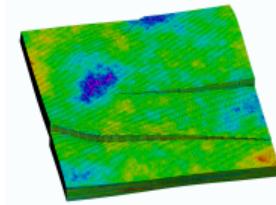


Res. -2



$14.84 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 97.11 \text{ dB}$

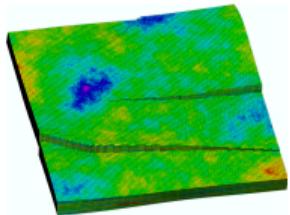
$21 \text{ MSB}_Z$



## Simulation: accuracy versus precision

Simulation (subjective)

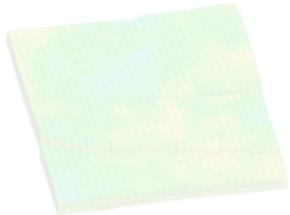
Res. 0



Res. -1

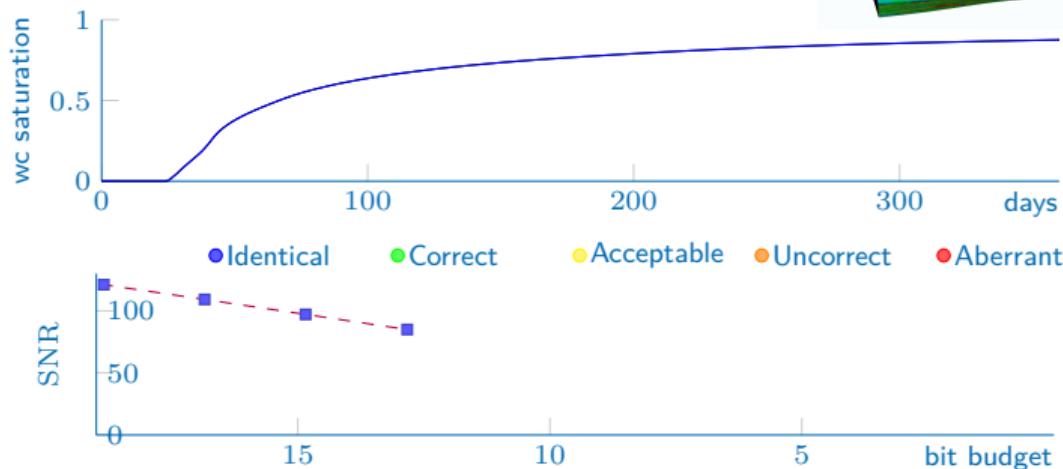
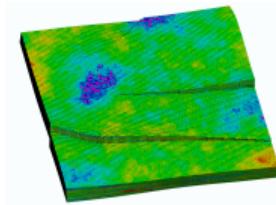


Res. -2



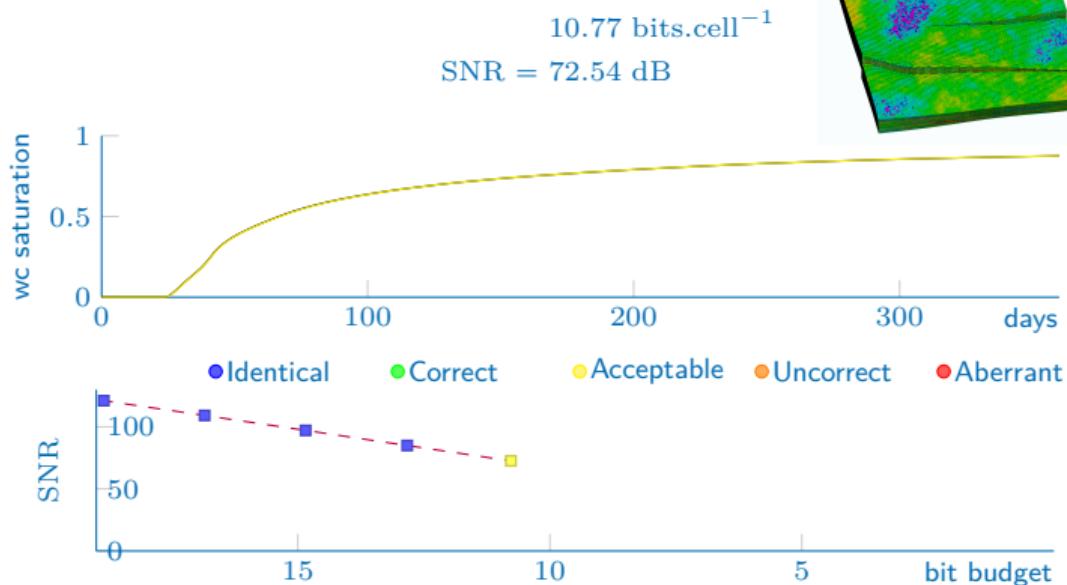
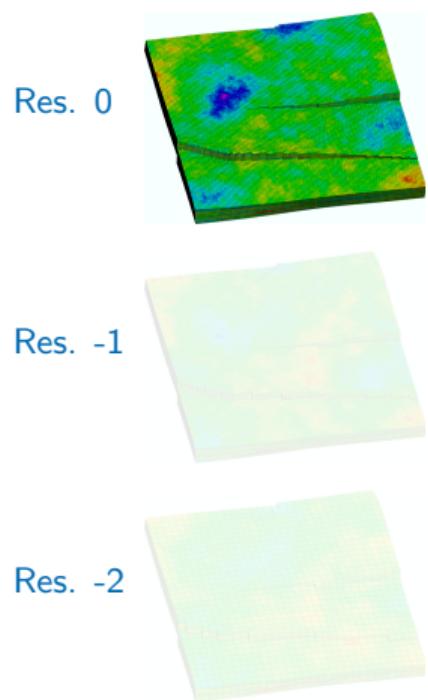
$12.83 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 84.94 \text{ dB}$

19 MSB<sub>Z</sub>



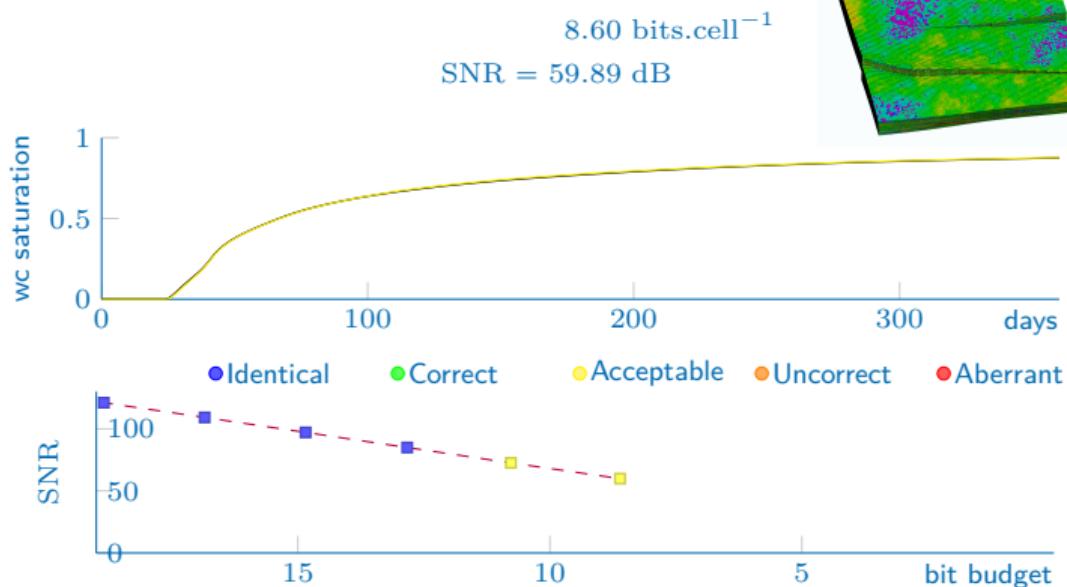
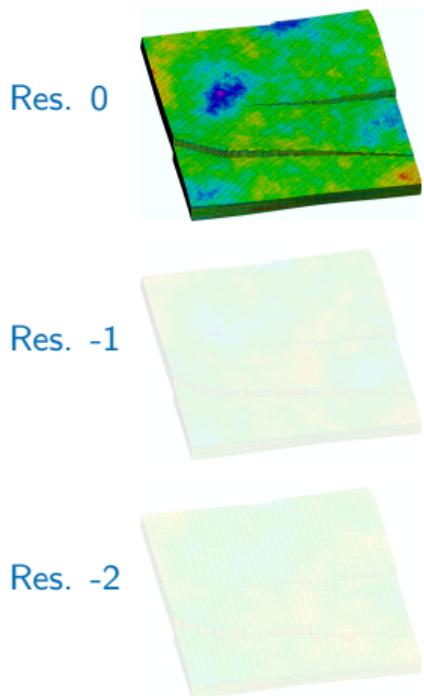
## Simulation: accuracy versus precision

Simulation (subjective)



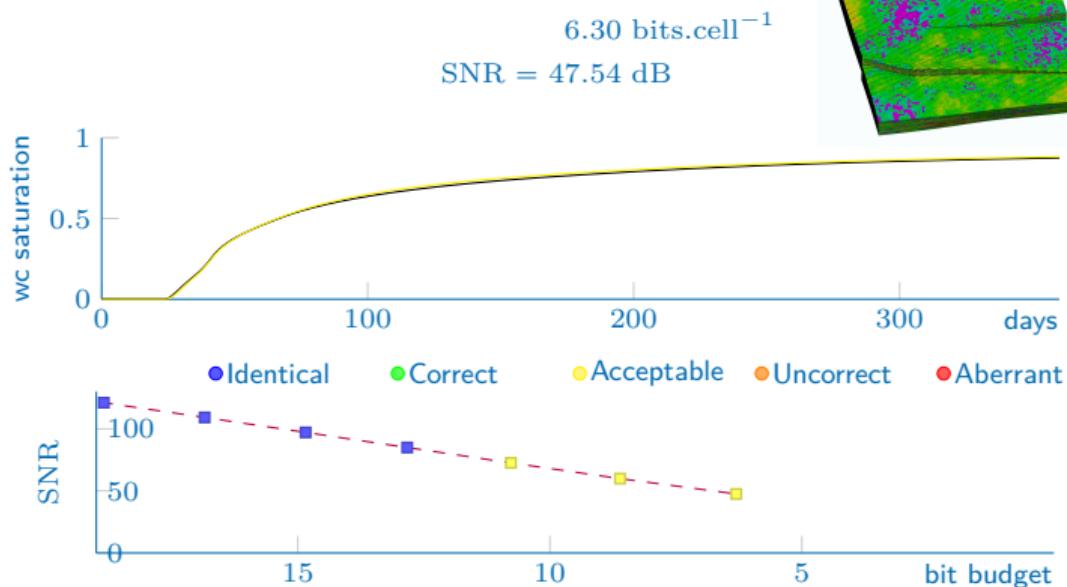
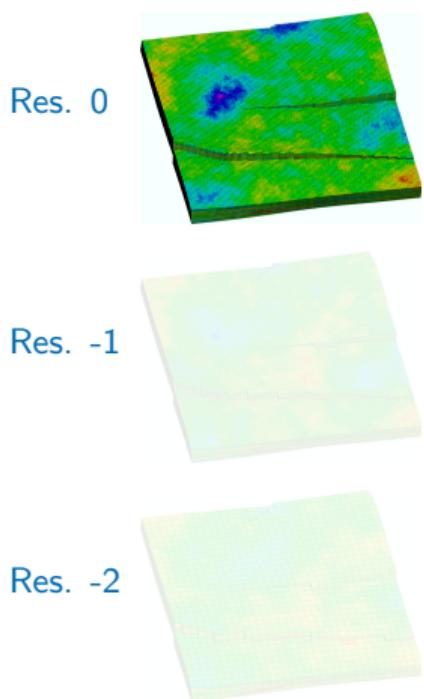
## Simulation: accuracy versus precision

Simulation (subjective)



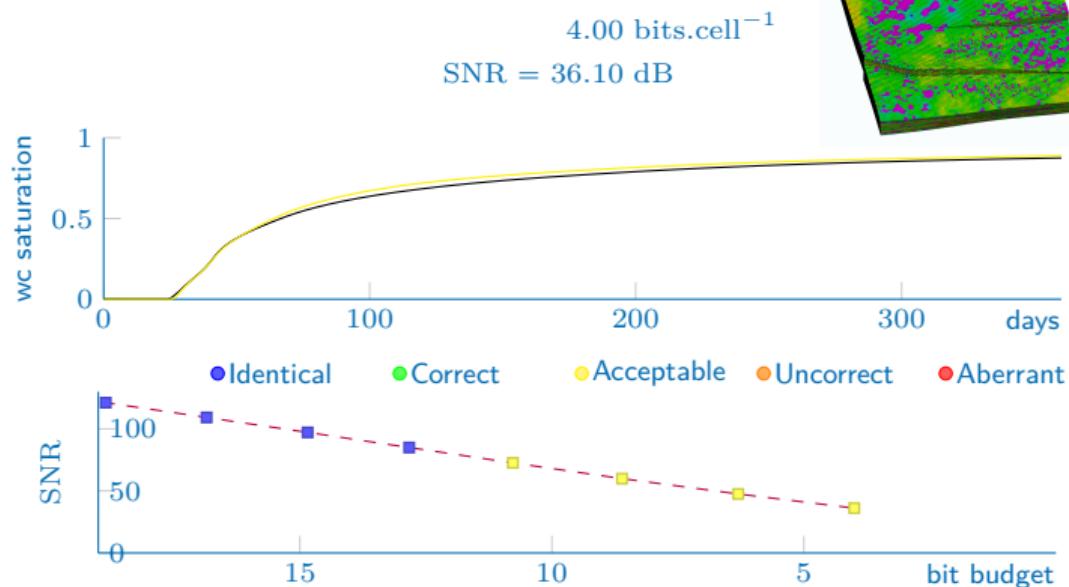
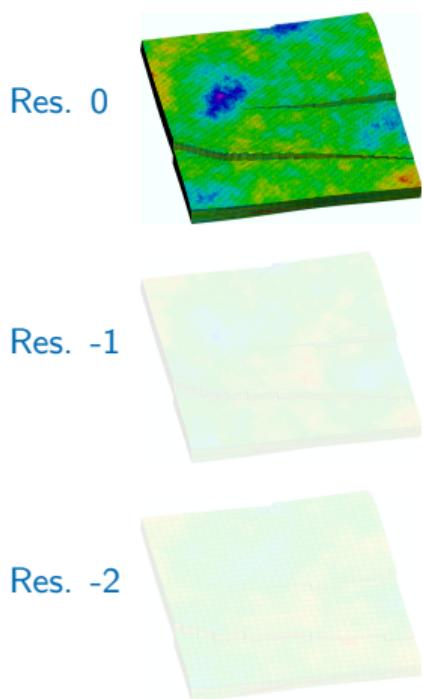
## Simulation: accuracy versus precision

Simulation (subjective)



## Simulation: accuracy versus precision

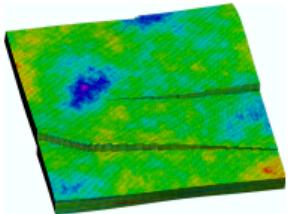
Simulation (subjective)



## Simulation: accuracy versus precision

Simulation (subjective)

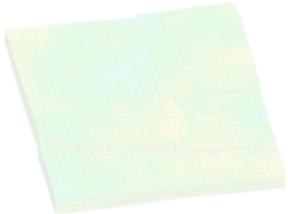
Res. 0



Res. -1

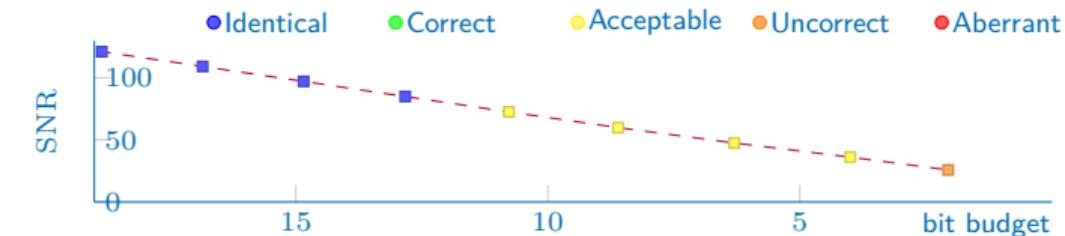
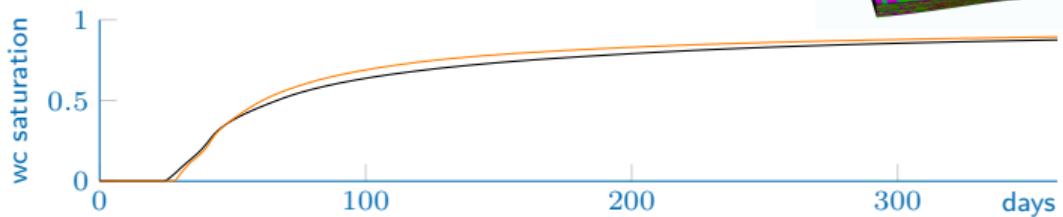
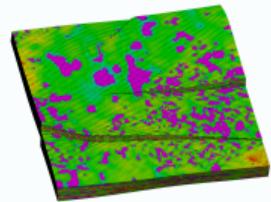


Res. -2



$2.05 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 25.82 \text{ dB}$

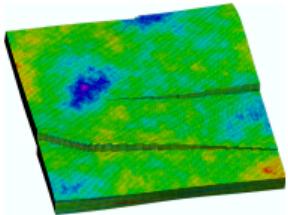
9 MSB<sub>Z</sub>



## Simulation: accuracy versus precision

Simulation (subjective)

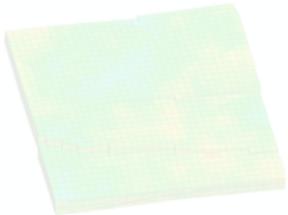
Res. 0



Res. -1

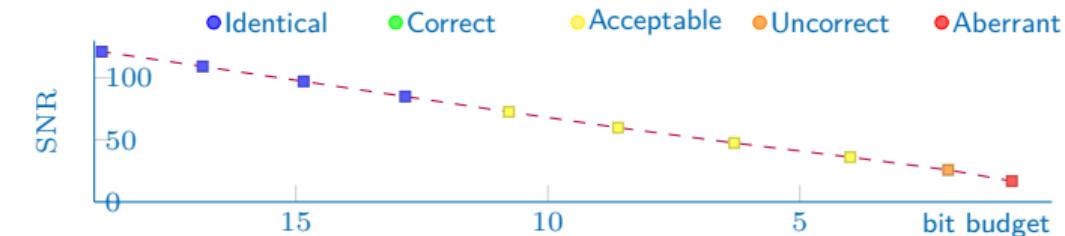
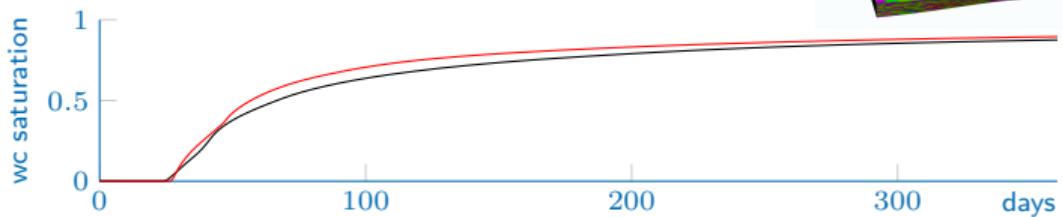
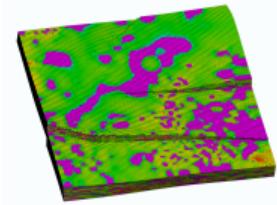


Res. -2



$0.79 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 16.91 \text{ dB}$

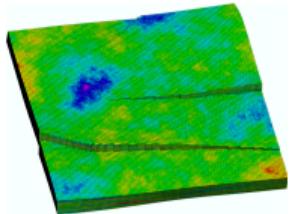
7 MSB<sub>Z</sub>



# Simulation: accuracy versus precision

Simulation (subjective)

Res. 0

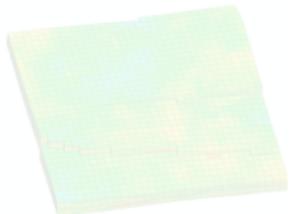


Is SNR suitable for simulation?

Res. -1



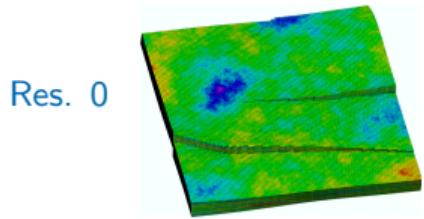
Res. -2



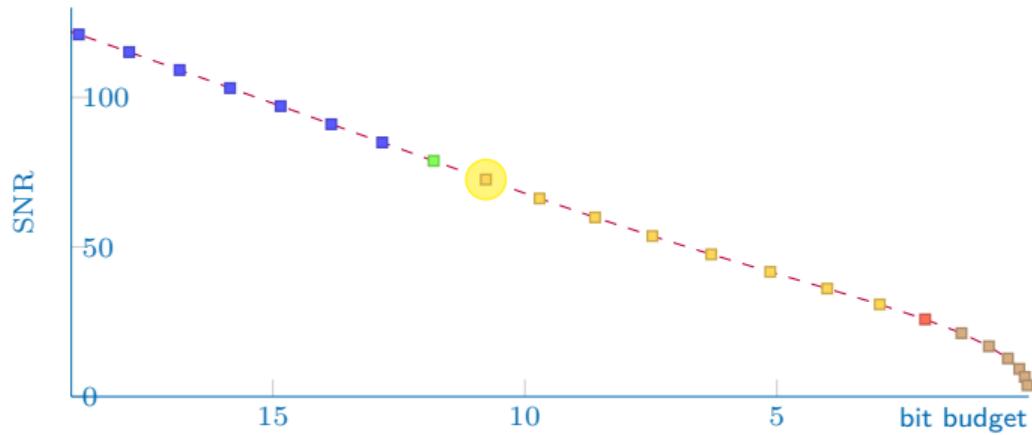
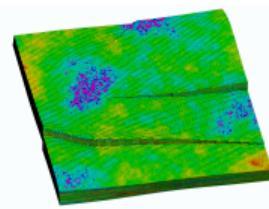
## Simulation-adapted metrics: is SNR valid?

$$\text{SNR} = 10\log_{10} \left( \frac{\sum_{c=1}^C P(c)^2}{\sum_{c=1}^C (P(c) - \hat{P}(c))^2} \right)$$

Simulation (better adapted metric)



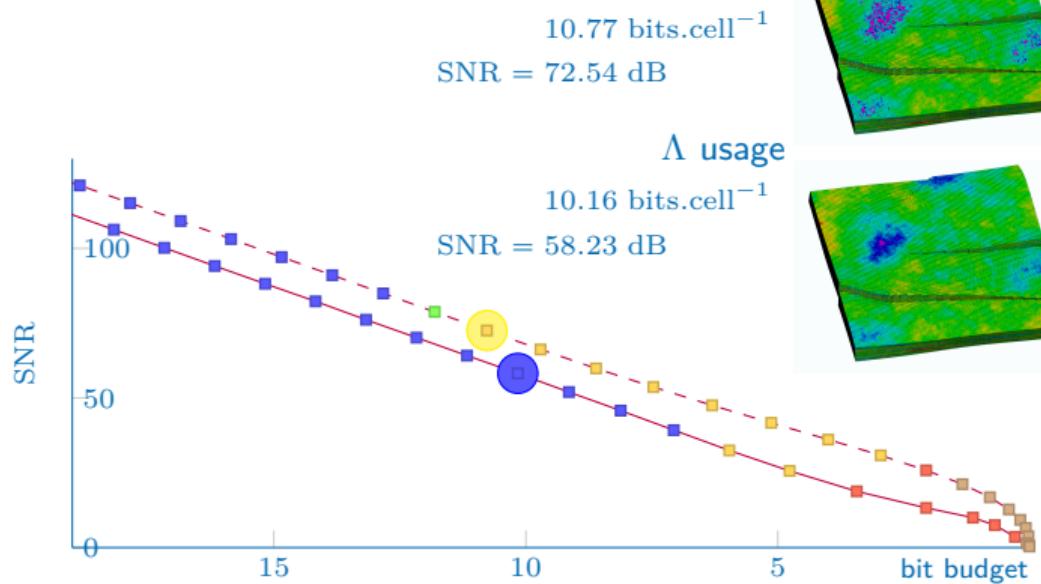
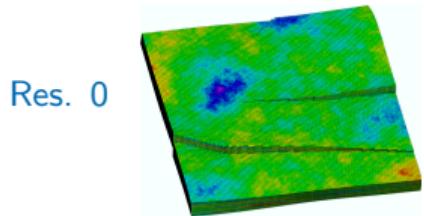
$10.77 \text{ bits.cell}^{-1}$   
 $\text{SNR} = 72.54 \text{ dB}$



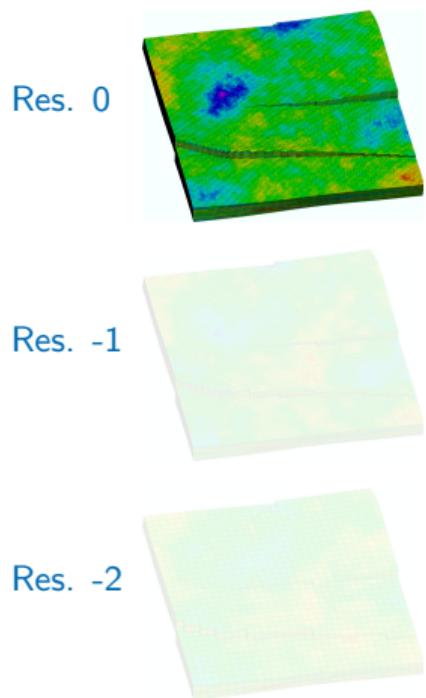
## Simulation-adapted metrics: is SNR valid?

$$\text{SNR} = 10\log_{10} \left( \frac{\sum_{c=1}^C P(c)^2}{\sum_{c=1}^C (P(c) - \hat{P}(c))^2} \right)$$

Simulation (better adapted metric)



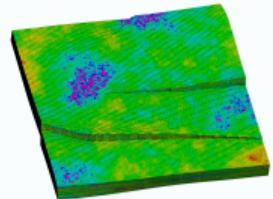
## Simulation-adapted metrics: is SNR valid?



$$\Lambda\text{-SNR} = 10\log_{10} \left( \frac{\sum_{c=1}^C \Lambda_{P(c)}^2}{\sum_{c=1}^C (\Lambda_{P(c)} - \Lambda_{\hat{P}(c)})^2} \right)$$

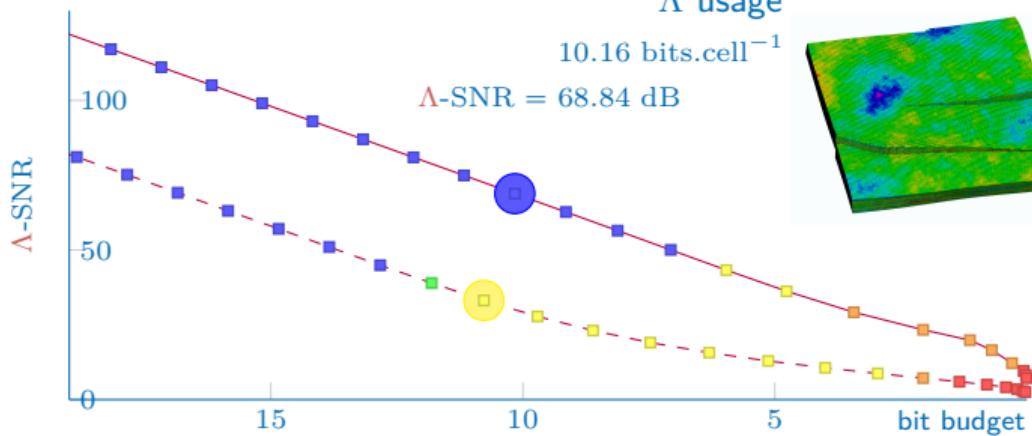
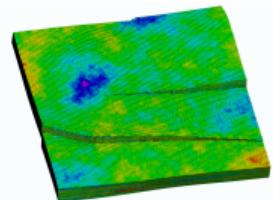
Simulation (better adapted metric)

$10.77 \text{ bits.cell}^{-1}$   
 $\Lambda\text{-SNR} = 33.14 \text{ dB}$



$\Lambda$  usage

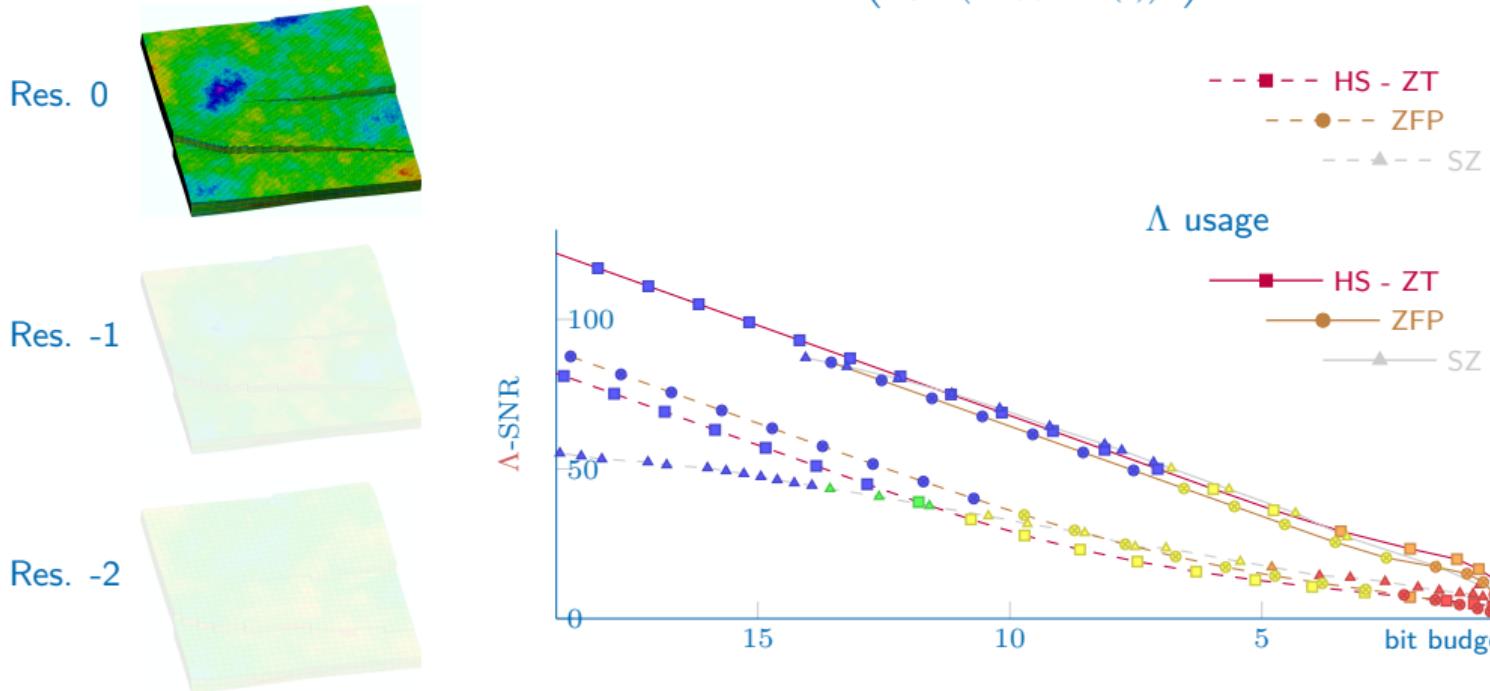
$10.16 \text{ bits.cell}^{-1}$   
 $\Lambda\text{-SNR} = 68.84 \text{ dB}$



## Simulation-adapted metrics: is SNR valid?

Simulation (better adapted metric)

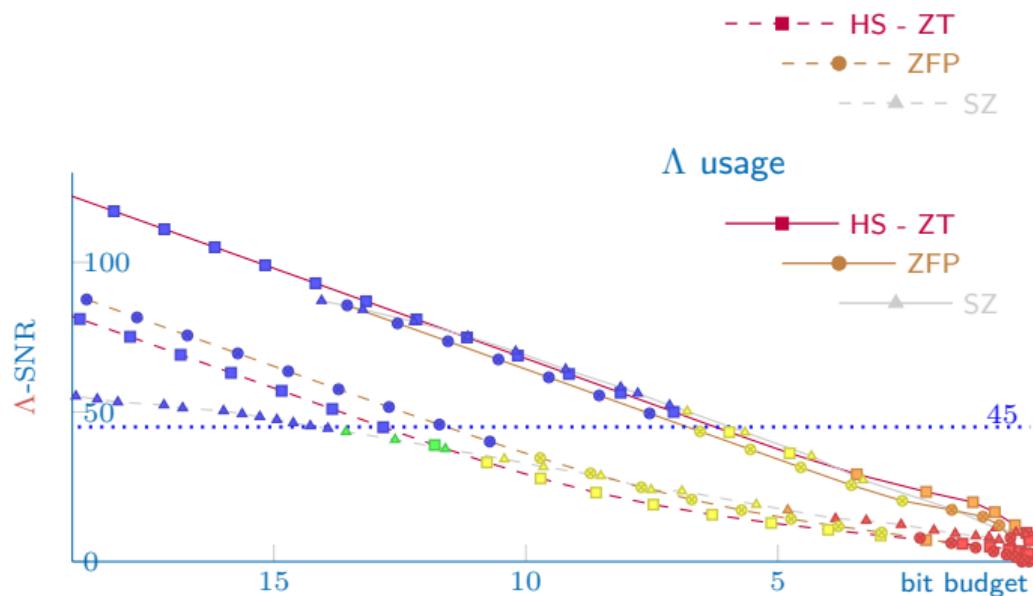
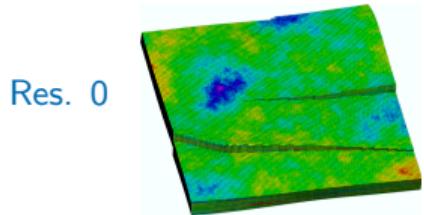
$$\Delta\text{-SNR} = 10\log_{10} \left( \frac{\sum_{c=1}^C \Delta_{P(c)}^2}{\sum_{c=1}^C (\Delta_{P(c)} - \Delta_{\hat{P}(c)})^2} \right)$$



## Simulation-adapted metrics: is SNR valid?

Simulation (better adapted metric)

$$\Lambda\text{-SNR} = 10\log_{10} \left( \frac{\sum_{c=1}^C \Lambda_{P(c)}^2}{\sum_{c=1}^C (\Lambda_{P(c)} - \Lambda_{\hat{P}(c)})^2} \right)$$



### Demanding domains

- Compression close to the workflow
- Integrate physical laws or proxies
- Develop pertinent metrics
- Require solid benchmarking

### Different needs

- Compress once, decompress many
- Data and energy efficient algorithms
- Random access: by the slice or by the block
- and many more

A lot of opportunities for DSP/compression people (from digital twins to Metaverse)

Visualization, storage, transfer, memory footprint, checkpointing restart, *in situ* analysis...

Energetic, economic, ecological

Some people in simulation

may not know well what compression can do (mesh inflation)

may have excessive expectation when they know more (all mesh random access)

Toward CaaS: Compression as a Service

Not the best in end-to-end quality, adapted to the overall workflow

With added value: old and new references

Natarajan, B. K., Filtering Random Noise from Deterministic Simals via Data Compression, 1995

Sabeti, E., Song, P. X. K., Hero, A. O., Data Discovery Using Lossless Compression-Based Sparse Representation, 2021

## Some teams to follow

Challenges

Allison H. Baker, National Center for Atmospheric Research (NCAR)

Franck Cappello, Argonne National Laboratory

Peter Lindstrom, Lawrence Livermore National Laboratory (LLNL)

Mark Ainsworth, Oak Ridge National Laboratory

And us? D, Bouard, Peyrot, Antonini, Payan