Compression Analysis

Preliminary Status Report

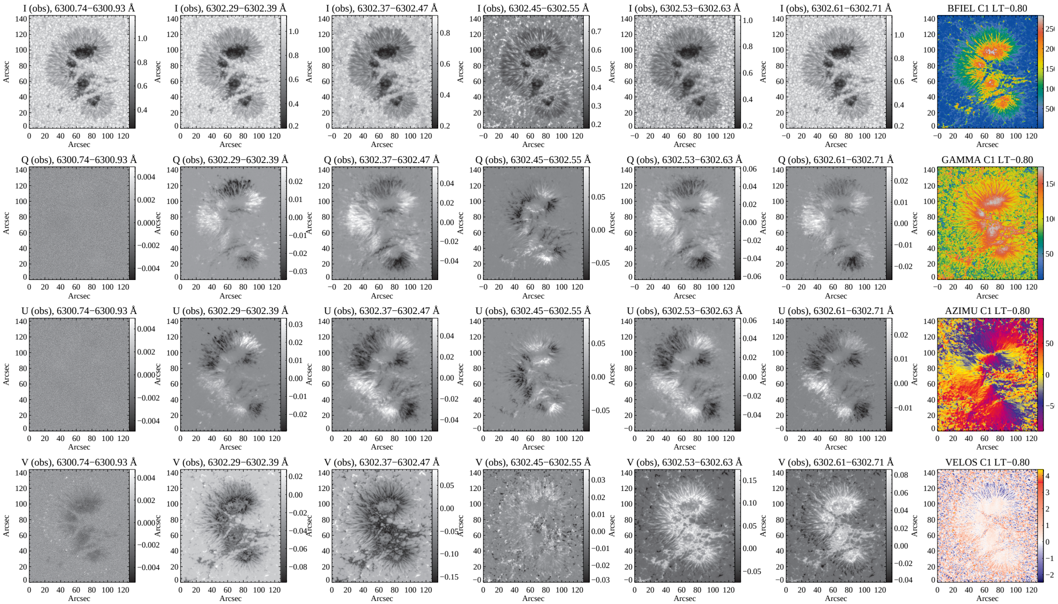
Andreas Lagg

Max-Planck-Institut für Sonnensystemforschung
Göttingen, Germany

ESAC, Jul 7-10 2014

Table of Contents

“ME-Compression”



Milne-Eddington "Compression"

ME-inversions: Pros

- Raw-data combined to 24 Stokes images (24 bit)
- on-board inversion: reduced to 1 – 5 images, 8 – 10 bit

→ extremely efficient LOSSY compression: factor 10 – 50

- almost no loss of science (ideal case only)

→ avoids errors in inversion caused by compression artefacts

ME-inversions: Cons

- very risky:

→ requires perfect on-board data reduction (dark images, flat fielding, fringe removal, polarimetric calibration, cross-talk removal)

- very complex: Sunrise II IMaX experience

Most sophisticated realization ever for on-board data volume reduction!

Milne-Eddington "Compression"

ME-inversions: Pros

- Raw-data combined to 24 Stokes images (24 bit)
- on-board inversion: reduced to 1 – 5 images, 8 – 10 bit

→ extremely efficient LOSSY compression: factor 10 – 50

- almost no loss of science (ideal case only)

→ avoids errors in inversion caused by compression artefacts

ME-inversions: Cons

- very risky:

→ requires perfect on-board data reduction (dark images, flat fielding, fringe removal, polarimetric calibration, cross-talk removal)

- very complex: Sunrise II IMaX experience

Most sophisticated realization ever for on-board data volume reduction!

Milne-Eddington "Compression"

ME-inversions: Pros

- Raw-data combined to 24 Stokes images (24 bit)
- on-board inversion: reduced to 1 – 5 images, 8 – 10 bit

→ extremely efficient LOSSY compression: factor 10 – 50

- almost no loss of science (ideal case only)

→ avoids errors in inversion caused by compression artefacts

ME-inversions: Cons

- very risky:

→ requires perfect on-board data reduction (dark images, flat fielding, fringe removal, polarimetric calibration, cross-talk removal)

- very complex: Sunrise II IMaX experience

Most sophisticated realization ever for on-board data volume reduction!

Lossy raw-data / Stokes-Vector Compression

(1) Raw data Compr.
(2) Stokes Vector Comp.

- allows to perform more sophisticated data analysis on ground
  (e.g. coupled inversions, height dependent atmospheres)
- allows better control on data reduction
  (fringe removal, PSF deconvolution, polarimetric calibration)
- Lossy compression noise must be below data noise
  (goal: 10^{-3})
- Compression ratios much lower than ME-compression

(3) Lossy compression of ME results

(3) Compression of ME parameter maps

- acceptable compression noise level higher than for Stokes data (0.5%, TBD)

$\longrightarrow$ Higher compression factors possible

- Influence on helioseismology data: to be investigated
- Influence on field extrapolations: to be investigated
- Achievable compression factors: TBD (science driven) $\longrightarrow$ SOPHISM simulations required
- Azimuth Ambiguity: lossy compression may lead to wrong results (to be investigated)

Stokes Vector Compression

Procedure

1. Use IMaX L2, Stokes *V* image (+40mÅ from line core)
2. reduce BPP
3. compress image with given ratio
4. decompress image

$\longrightarrow$ determine file size ratio

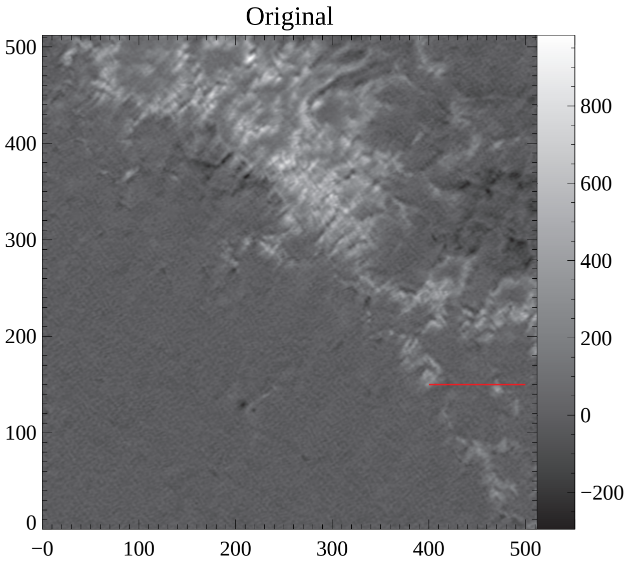
$\longrightarrow$ compute difference: map, RMS

Compression Code

- provided by IDA Braunschweig
- several lossy algorithms available
- same code will run on-board (FPGA)
- implemented to SOPHISM (June 2014)

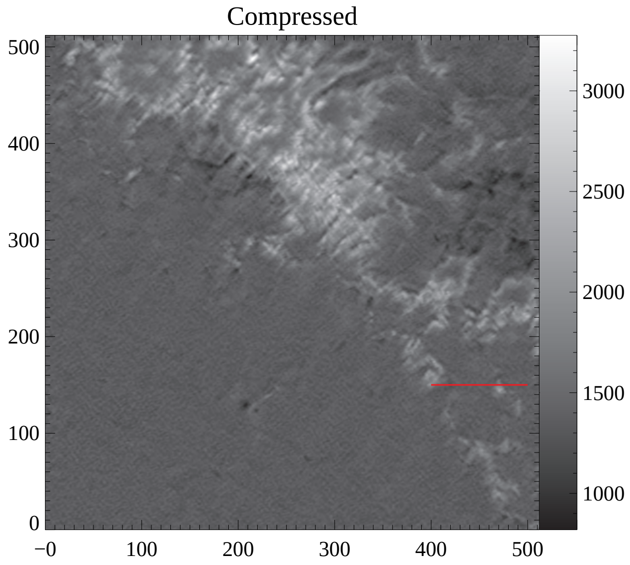
IMaX Stokes V

Original



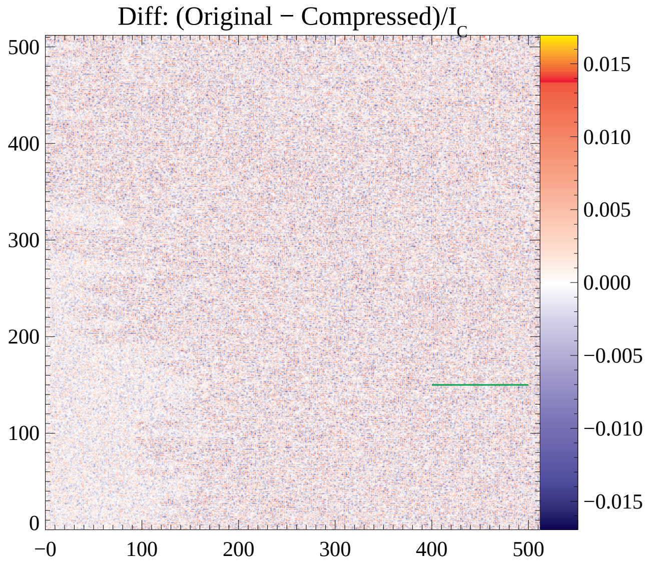
IMaX Stokes V

Compressed:
12 BPP, ratio: 8.0
size reduction
$\times$10.67,
RMS=3.02E-03



IMaX Stokes V

Original - Compressed
12 BPP, ratio: 8.0
size reduction ×10.67,
RMS=3.02E-03



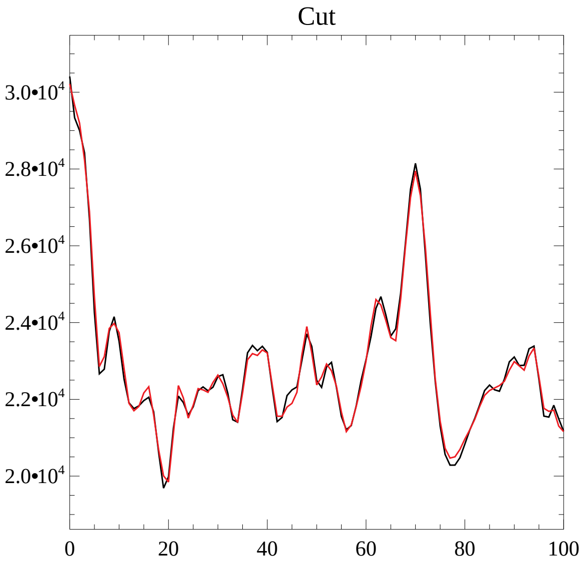
IMaX Stokes V

Cut: Original, Compressed

12 BPP, ratio: 8.0

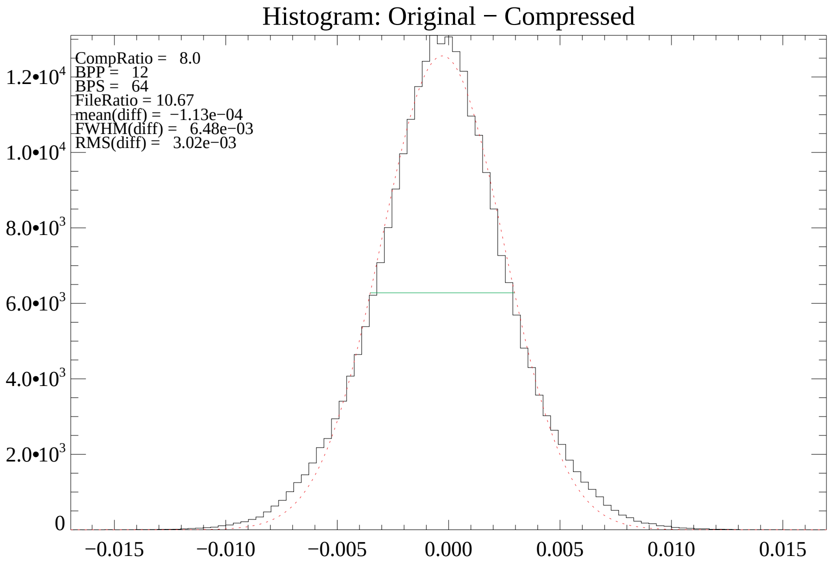
size reduction $\times$10.67,

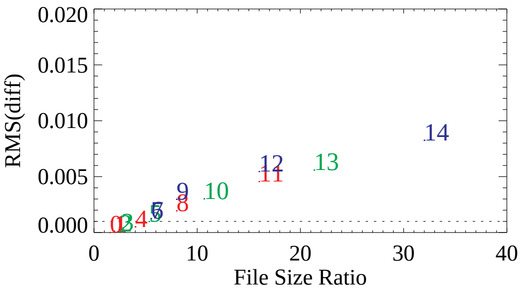
RMS=3.02E-03

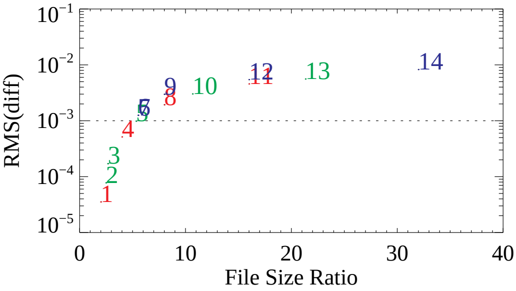


IMaX Stokes V

Original - Compressed
12 BPP, ratio: 8.0
size reduction $\times$10.67,
RMS=3.02E-03







SUMMARY: StokesV, IMaXL2

	FileRatio	CompRatio	BPP	BPS	mean(diff)	FWHM(diff)	RMS(diff)
0	1.54	1.0	16	64	0.00e+00	0.00e+00	0.00e+00
1	2.00	2.0	16	64	-2.91e-06	8.96e-05	3.53e-05
2	2.49	1.0	12	64	-1.49e-08	2.23e-04	7.75e-05
3	2.67	2.0	12	64	-5.48e-06	2.77e-04	1.71e-04
4	4.00	4.0	16	64	-6.97e-06	9.51e-04	5.13e-04
5	5.33	4.0	12	64	-1.32e-04	2.09e-03	9.86e-04
6	5.53	1.0	8	64	3.24e-08	3.56e-03	1.25e-03
7	5.53	2.0	8	64	3.24e-08	3.56e-03	1.25e-03
8	8.00	8.0	16	64	-2.11e-06	4.43e-03	1.94e-03
9	8.00	4.0	8	64	-9.41e-05	5.56e-03	2.98e-03
10	10.67	8.0	12	64	-1.13e-04	6.48e-03	3.02e-03
11	16.00	16.0	16	64	-6.90e-07	9.36e-03	4.56e-03
12	16.00	8.0	8	64	2.93e-04	1.19e-02	5.46e-03
13	21.33	16.0	12	64	-2.52e-05	1.23e-02	5.59e-03
14	32.00	16.0	8	64	7.92e-04	1.86e-02	8.25e-03

Summary - Stokes Vector Compression

Lossless Compression

Compression factors: $\times 1.4 - \times 1.9$

BPP Compression

noise level for bit truncation always higher than lossy compression at same file size ratio

Lossy Compression

achievable compression ratios for compression noise levels of $\leq 10^{-3}$ (Stokes QUV, all raw-data images)

$\longrightarrow \approx \times 4 - \times 5$

for compression noise levels of 0.5% (maybe sufficient for Stokes I)

$\longrightarrow \approx \times 10$

Preliminary Results! SOPHISM analysis still to be done!

Summary - Stokes Vector Compression

Lossless Compression

Compression factors: $\times 1.4 - \times 1.9$

BPP Compression

noise level for bit truncation always higher than lossy compression at same file size ratio

Lossy Compression

achievable compression ratios for compression noise levels of $\leq 10^{-3}$ (Stokes QUV, all raw-data images)

$\longrightarrow \approx \times 4 - \times 5$

for compression noise levels of 0.5% (maybe sufficient for Stokes I)

$\longrightarrow \approx \times 10$

Preliminary Results! SOPHISM analysis still to be done!

Further Compression Tests

Critical Issues:

- helioseismology / extrapolation aspect of compression
- bad pixel treatment
- vector magnetic field maps: Compression without disambiguation possible?

New Test Setup:

- use SOPHISM
- test full pipeline

→ only then results have the necessary level of realism

Test Matrix

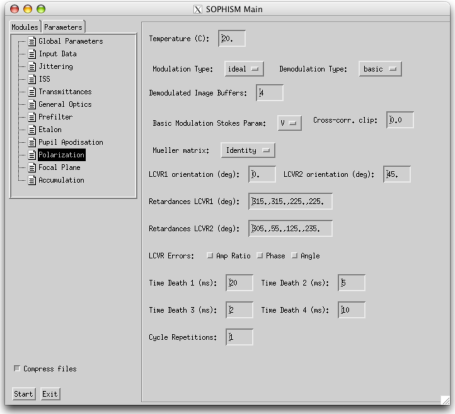
- raw data, Stokes vectors, ME-maps
- FDT, HRT
- lossless, $\times$2,4,8,12
- bit depth variation: 16,14,12,10
- compression algorithms
- data with 2 different noise levels

Total: $\approx$ 800 tests

Time Table:

- computations: end of September
- results: end of October

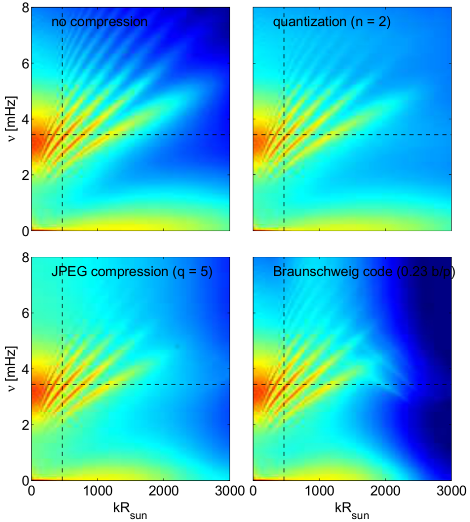
SolO PHI Instrument Simulator (SOPHISM)



(Julian Blanco, Univ. Valencia)

- Software IDL-coded SO/PHI simulator
- Main effects simulated in quasi-independent modules:
  - Jittering+ISS
  - Polarization modulation
  - Spectral profile
  - Optical aberrations
  - Pupil apodization
  - FPA, Accumulation
  - Modulation scheme
  - Solar evolution
  - Compression
  - ....
- Modules generate result files

Helioseismic Power Spectra

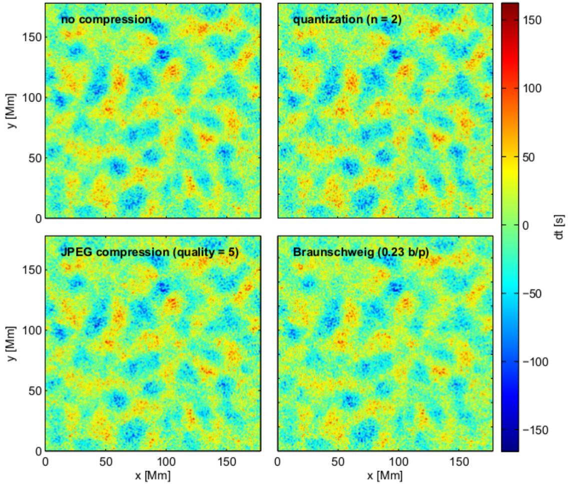


Source: v_{LOS} maps

- Compression barely affects waves
- Quantization: almost constant background noise
- JPEG and Braunschweig code:
  - Artifacts in power spectrum
  - Granulation reduced (important for LCT)

(Björn Löptien, MPS)

Travel-time Maps of Supergranulation

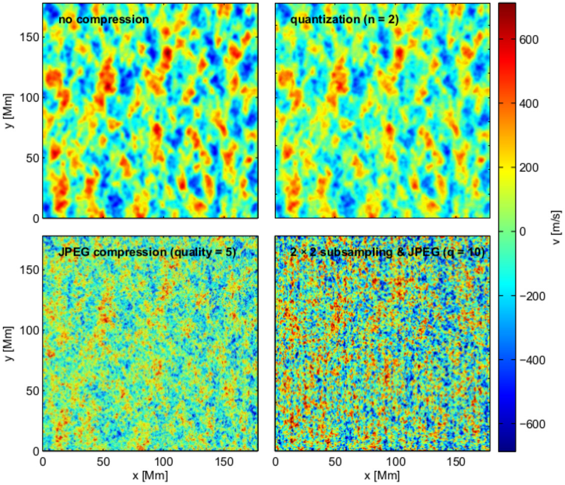


Source: v_{LOS} maps

- Divergence of supergranulation flows
- F-mode (near-surface)
  Negative: outflows
  Positive: inflows
- Maps from compressed data almost indistinguishable from uncompressed data No apparent correlation between noise and travel-times

(Björn Löptien, MPS)

LCT of Supergranulation



Source: I_C maps

- v_x on supergranulation scale
- Strong influence of JPEG compression and Braunschweig code

(Björn Löptien, MPS)

Summary Helioseismology (PRELIMINARY!)

- Only one simple measurement tested so far (f-mode helioseismology of supergranulation)
- Results probably not valid for helioseismology in general
- Tolerable compression efficiency depends on measurement method and science goal
- Time-distance: JPEG best method that we tested so far
- LCT: Quantization best method that we tested so far

Next steps:

- Test on near-surface latitudinal differential rotation (ongoing work)
- Test on HMI raw data (center-to-limb effects, e.g., foreshortening)

(Björn Löptien, MPS)