Compression Algorithms for High Data Volume Instruments on Planetary Missions: a Case Study for the Cassini Mission

Hua Xie, Robert West, Benoît Seignovert, Jeffrey Jewell, William Kurth, Terrance Averkamp

Compression Algorithms for High Data Volume Instruments on Planetary Missions: a Case Study for the Cassini Mission

Hua Xie[a,*], **Robert A West**[b], **Benoît Seignovert**[b], **Jeffrey Jewell**[c], **William Kurth**[d], **Terrance Averkamp**[d]

[a]Jet Propulsion Laboratory, California Institute of Technology - Communications Architectures and Research Section.
[b]Jet Propulsion Laboratory, California Institute of Technology - Planetary Science Section.
[c]Jet Propulsion Laboratory, California Institute of Technology - Instrument Software and Science Data Systems Section.
[d]Department of Physics and Astronomy, University of Iowa.

Abstract. We investigated data compression algorithms to boost science data return from high data volume instruments on planetary missions, particularly outer solar system missions where every bit of data represents an engineering triumph of over severe constraints on mass (limiting antenna size) and power (limiting signal strength). We developed a methodology to (1) investigate algorithms to improve compression, and (2) to work with the science teams to evaluate the effects on the science.

Our algorithm for compressing the Cassini RPWS (Radio Plasma Wave Science) data achieved a factor of 5 improvement in data compression (relative to what the RPWS team was using), and our algorithm for the Cassini UVIS (Ultraviolet Imaging Spectrograph) Saturn data set achieved a much higher factor ($\sim$70). In both cases, the investigators on the science teams who evaluated our results reported that the science goals were not compromised. Our compression algorithm for ISS (Imaging Science Subsystem) images achieved on average a factor of $\sim$1.7 improvement in lossless compression compared to the original algorithm. We also evaluated the compression effectiveness of JPL's Fast Lossless EXtended (FLEX) hyperspectral/multispectral image compressor on Cassini's VIMS (Visible and Infrared Mapping Spectrometer) data. FLEX lossless compression provides a factor 2 improvement over the original compression. We also explore a different range of lossy compression which can achieve an additional factor 2 to 5 depending on the fidelity required.

Our findings have implications for the design of future space missions, particularly with respect to antenna size and overall Size, Weight, and Power (SWaP) budgets, by demonstrating strategies to implement better data compression. In addition to improved algorithms, we show that an iterative process involving real-time science team evaluation and feedback to update the on-board compression algorithm is both essential and feasible. We make the case that a spacecraft facility compressor hosting a toolbox of compression algorithms, available to all of the science instruments and supported by a team of compression experts, conveys significant benefits. Beyond the obvious benefits of increased science return and faster playback, better data compression enables design trades between antenna size and number of science instruments on the payload.

Keywords: Data compression, Solar system exploration, Sparse coding, Deep space communications.

***Corresponding author:** hua.xie@jpl.nasa.gov

1 Introduction

Given the tension between science return and cost of interplanetary missions, a fresh look at lossy compression methods and a careful evaluation of what level of error is tolerable would be in the interest of higher science return and/or lower mission cost. This is especially true for missions to the outer solar system (Jupiter and beyond) where downlink rates are highly constrained. Current practice for data compression largely follows the historical paradigm whereby each instrument on a spacecraft payload provides, or not, its own compressor. The hardware and software for compression are built well ahead of launch. This practice leaves little ability to alter compression schemes based on lessons learned in the early part of the mission. Consequently, the amount of data returned is always less than what could be achieved with the 'optimal' compression scheme.

The science can suffer either because the compressor produces unacceptable artifacts (the Cassini Imaging Science Subsystem, ISS, is one example) or because the compression strategy is overly conservative (e.g. the Cassini Visual and Infrared Mapping Spectrometer, VIMS, did not have the ability to do lossy compression).

Our intent is not to recommend the 'best' compressor - something that depends on the instrument, the mission, and science/data volume tradeoffs that only the science teams can judge. Rather, we illustrate the issues more fully using examples from several instruments on the Cassini mission, and we build a case for changing the paradigm of how data compressors might be implemented for future missions. The key idea is to include on the spacecraft a hardware and software compression toolbox with a variety of efficient compression algorithms, and with the ability to add to and make changes during flight, to the algorithms and to the parameters that govern the compression. With these capabilities the science teams have the opportunity to optimize compression, taking advantage of hindsight gained from the early part of the mission, and using tools that could be designed or selected by the science teams and compression experts prior to launch. By bringing in compression experts at the mission level takes some of the burden off of the science team to come up with the best compression strategy. The examples we give illustrate how this would work if we had the ability to make changes or to add new algorithms during the Cassini mission.

The examples that follow simulate how this would work on a hypothetical mission that is similar to Cassini, but with the ability of the science team to evaluate compression during the early phase of the mission, and to make changes or to add new algorithms during the mission. It is not our goal to recommend a 'best' algorithm, and there may be algorithms and hardware compressors that do a better job than the ones we selected for the illustrations. The algorithms we use for the illustration are ones that are readily available to us and, as it turns out, perform better than the ones used on the Cassini mission.

A key issue for lossy compression is the risk of science loss. This concern is a driver for conservatism that contributes to under-utilization of lossy compression and it must be addressed in the context of our study. Therefore we have included science evaluation in each of our examples. Science evaluation in real time would be required as part of a new mission to implement the approach we recommend. Ideally science teams would evaluate, during the mission, how well the science holds up under various algorithms or choices for compression parameters. In this regard, we performed such an evaluation for a small part of the Cassini RPWS (Radio Plasma Wave Science) data set where the science (the detection of waves and dust particle impacts) could be easily assessed. However, for the other data sets we opted to compare lossy compression errors against instrumental noise. The rationale being that, if the compression errors are smaller than the instrumental noise it is likely that the science loss would be acceptably small. In practice it would be up to the science teams to come up with metrics for what loss is acceptable.

Our ultimate goal is to demonstrate compression algorithms for future outer solar system space missions able to directly incorporate scientifically motivated distortion metrics in the compression scheme. This will enable us to develop a framework to generalize the classical Shannon rate-distortion theory to optimize scientific retrieval performance of the reproduced data. Exploring the ideas above for science motivated distortion metrics was beyond the scope of this paper, as it involves the computationally demanding merger of detailed forward physical models and optimal estimation with compression. Future work will endeavor to do just that.

The article is organized as follows. Section 2 provides an overview of the Cassini mission, its instrumentation, and onboard compression algorithms. Section 3 describes the suite of algorithms

we investigated for more effective compression of Cassini's high data volume instrumental data . Section 4 introduces an iterative process we established to update and improve the compression using science team evaluation and feedback. Section 5 reports the results of applying these compression algorithms on the Cassini data sets and discusses the impacts they had on science quality. To lead to future development of science-driven data compression methodology, we provide in Section 6 an example of scientific distortion metric for spectrometer data. We envision that this type of distortion metric, which involves the detailed physical model of the underlying process, can be incorporated in data compression optimization via machine learning methods. Section 7 concludes this paper and discusses future work.

2 Overview of Cassini, Its Instrumentation, and Compression Algorithms

To illustrate the concepts we wish to explore we use several examples from the Cassini mission which flew by Jupiter and orbited Saturn from 2004 to 2017. We chose this mission because we are familiar with data from several high-data-volume instruments, including some of the associated science issues. We carried out a simulation to answer the question of 'How could science return from these instruments be improved if we had the ability to adjust the types of algorithms used for data compression, based on a fraction of the data received from the early part of the mission?'. A boundary condition for this exercise was to impose limits on compression ratios consistent with no or negligible loss of science. By illustrating how this would be done for Cassini instruments we can demonstrate and recommend procedures to be applied to future orbital or long-duration missions.

Our selection of instruments serves to illustrate what algorithms and science evaluations could be carried out for one-dimensional time-series data (the Cassini RPWS experiment), two-dimensional images (the ISS) and three dimensional multi-spectral images (1-spectral by 2-spatial; Cassini VIMS and UVIS). The last two examples differ greatly in terms of signal/noise ratio, with profound implications for data compressibility.

The Cassini/Huygens mission was a NASA/ESA Flagship mission whose principal targets were the components of the Saturn system (planet, rings, moons, magnetosphere). Saturn's moon Titan provided special interest, and the mission included a probe that descended into Titan's atmosphere and landed on the surface. Here we consider four instruments on the orbiter, the RPWS, ISS, UVIS, and VIMS. Details on the science goals and design of the experiments are given by Porco et al.[1] for ISS, Gurnett et al.[2] for RPWS, Brown et al.[3] for VIMS, and Esposito et al.[4] for UVIS.

The Cassini RPWS instrument was designed to study radio and plasma waves in the Saturnian system along with thermal plasma properties. Its seven sensors comprise three nearly orthogonal electric antennas, each 10 m in length, three orthogonal search coil magnetometers, and a Langmuir probe. A set of five receivers sampled the spectrum from 1 Hz to 16 MHz for electric fields and from 1 Hz to 12 kHz for magnetic fields. The instrument could return both spectral data as well as time-series data. The time series burst-mode data comprised electric and/or magnetic field intensities from one or more of the sensors with sample rates ranging from 100 12-bit samples/s (sps) to 222,222 8-bit sps (222.2 ksps). In this article we used the data obtained by the Cassini WBR at $75\,\mathrm{kHz}$ filter bandwidth. These data sources had to be managed throughout the mission; the telemetry devoted to them was always a subject of negotiation with other science teams. The RPWS instrument could limit its data production by limiting the times of collecting burst data, through duty cycling, and through data compression. The onboard compression was restricted to

lossless using either a Rice compression scheme or a Walsh Transform followed by Rice compression.[5,6] The onboard compressor typically achieved compression ratios of $4:1$. For some RPWS time series with very high data entropy waveforms including impulsive spikes, the achieved compression ratio was close to $2:1$.

The ISS used two cameras, a Wide Angle Camera (returned 124,713 images) and a Narrow Angle Camera (318,358 images). Both used CCD detectors with format 1024×1024 pixels, and with 12 bits/pixel in default mode. The instruments could be commanded to convert the 12 bits to 8 bits according to a square-root lookup table. That operation preserved dynamic range but increased quantization uncertainty. Another option, rarely used, allowed for the return of the 8 least significant bits. In addition, there were options for lossless compression using Huffman encoding (but with loss of data for high-entropy targets), and lossy JPEG-type compression using a hardware compressor supplied by Matra-Marconi. The lossy compressor produced some unwanted artifacts and was not used in the latter half of the mission. In this work we make use of a small set of uncompressed images.

The UVIS included an EUV (Extreme Ultraviolet) imaging spectrograph, an FUV (Far Ultraviolet) spectrograph, a High Speed Photometer and a Hydrogen/Deuterium Absorption Cell. In this article we looked at Saturn spectral images taken with the FUV. The instrument used a 2-dimensional array with a microchannel plate cathode followed by a coded anode array, with 1024 spectral columns and 64 spatial rows. A second spatial dimension was obtained by building up spectra as the slit moved in a direction perpendicular to its long dimension, or by acquiring multiple spectra at fixed pointing as the planet rotated. In this study we examined more than 8,000 spectra taken in the fixed pointing mode. This data set comprised a small fraction of the total data with Saturn as the target. It made use of the longest available integration time (20 minutes per spectrum) to achieve maximum signal/noise ratio. The raw data consist of packets of 9-bit integers, reduced via a modified square-root procedure from up to 16 bits. The ground software inverts this operation. The quantization noise resulting from this operation does not significantly add to the photon shot noise. Although these data can be considered as a three-dimensional array, we used individual spectra to explore the Singular Value Decomposition compression method, unlike the case for the VIMS data set where the compression technique used for those data takes advantage of the full three-dimensional spectral image cube structure.

The VIMS investigation is a multidisciplinary study of the Saturnian system using visual and near-infrared imaging spectroscopy and high-speed spectrophotometry techniques. The scope of the investigation includes the rings, the surfaces of the icy satellites and Titan, and the atmospheres of Saturn and Titan. The VIMS was composed of 2 hyper-spectrometers co-aligned: VIMS-VIS observed the visible part of the spectrum in 96 channels ($0.35\,\mu\mathrm{m}$ to $1.04\,\mu\mathrm{m}$) and VIMS-IR covered the infrared range ($0.88\,\mu\mathrm{m}$ to $5.12\,\mu\mathrm{m}$) with 256 channels allowing a small overlap around $1\,\mu\mathrm{m}$.[3] Both instruments observed a slightly different field of view with a maximal extent of 64×64 pixels in the spatial dimension. In this article we investigated a small fraction of the VIMS cubes acquired during Titan flybys. Titan VIMS data helped to constrain gas and aerosol distributions in Titan's atmosphere as well as their variability with time. In addition, the VIMS-IR data also sees the surface of Titan through atmospheric windows which helps to determine surface properties. The VIMS data covered a wide variety of illuminations and many emisision angles. Depending on the distance between the spacecraft and the target, the VIMS measurments were operated in a wide range of spatial and spectral resolutions to meet the scientific objectives. Onboard the Cassini spacecraft, VIMS data were losslessly compressed using a preprocessor routine and Rice encoder.[7]

The preprocessing routine calculated the difference between the image line (64 pixel line) and a representative "brightness" line, and then Rice encoded the difference image into the compressed bit stream. The compression ratio ranged between $2:1$ and $3:1$ depending on the actual entropy of the data.

3 Data Compression Algorithms Investigated In This Study

Space-based sensor data tends to be unique for each mission and each instrument. Current practice for data compression largely follows the historical paradigm where each instrument on a spacecraft payload provides its own compressor. The hardware and software for compression are built well ahead of launch. This practice leaves little ability to alter compression schemes based on lessons learned in the early phase of the mission.In this retrospective study, we investigated a suite of compression algorithms for more effective compression of data from four high data volume instruments onboard the Cassini spacecraft. We looked at both the algorithms that are more able to compress data relative to those used by the Cassini instruments, and a methodology to address the concerns of the science teams regarding use of lossy compression.

Our intent is not to recommend the 'best' compressor - something that depends on the instrument, the mission, and science/data volume tradeoffs that only the science teams can judge. Rather, we illustrate the benefits of having the ability of the science team to evaluate compression during the early phase of the mission, and to make changes or to add new algorithms during the mission. The algorithms we use for the illustration are ones that are readily available to us and, as it turns out, perform better than the ones used on the Cassini mission.

The following sections describe the data compression techniques we selected for investigation in this study. We focused on transform coding and predictive coding methods because of their flight heritage and potential enhancements that can be incorporated into these schemes. We provided examples to demonstrate benefits of adjusting and enhancing the compression methods based on scientific evaluation feedback, and we build a case for changing the paradigm of how data compressors might be implemented for future missions.

3.1 Transform Coding

Transform coding methods have been used for data compression by many missions, e.g., Mars Pathfider,[8] Cassini ISS,[1] and Mars Exploration Rovers.[9] Transform compression removes spatial redundancy by first performing a linear (or nearly linear) transform designed to decorrelate the signal by local separation of spatial frequencies. The main objective of transform coding is that a good transform will yield uncorrelated components, concentrating the signal energy in a few of the coefficients, leading to lower entropy rate in the tranform domain. The coding step can quantize (lossy) and entropy code the transform coefficients based on a given rate constraint and distortion metric. At the receiver, the quantized transform coefficients can be combined with a set of basis functions to reconstruct the original (lossless) or an approximation (lossy) of the original signal. Transform coding can be either lossless and lossy depending on whether a quantization step is applied.

The Consultative Committee for Space Data Systems (CCSDS) has been engaging in recommending data compression standards for space applications since 1997. A Discrete Wavelet Transform (DWT) based method[10] is adopted as the CCSDS standard for on-board compression of 2D images. For example the ICER compressor[11] used by Mars Exploration Rovers, is a progressive,

wavelet-based image data compressor designed to meet the specialized needs of deep-space missions while achieving state-of-the-art compression effectiveness.

While existing standards for both commercial[12] and space applications[10] utilize wavelet transforms for image compression, in this study we also investigated alternative transforms, e.g., wave atoms[13] and SVD,[14] that can be more effective for representing and compressing the types of data investigated herein thus providing enhancements to existing transform coding methods.

In this article we investigated transform coding using wavelets[15] and wave atoms[13], for compression of RPWS time series and ISS images. Both wave atom and wavelet transforms belong to the wavelet packets multiresolution decomposition family introduced by Coifman.[16] Wavelet packets comprise the entire family of subband coded (tree) decompositions, where one applies a special pair (highpass and lowpass) of digital filters to the signal and then subsamples each by factor of two. This subband filtering is applied recursively to chosen frequency intervals, resulting in different partitioning of the time-frequency domain. Each partition represents a certain trade-off in time versus frequency resolution. One can not have arbitrarily fine time and frequency resolutions. Instead different transforms trade frequency resolution for time resolution and vice versa. For example, a short-time Fourier transform (Gabor transform) has constant frequency and time resolutions for all frequency components; a wavelet transform provides high time resolution for high frequency components, but has poor frequency resolution for them; a more advantageous compromise between time resolution and frequency resolution is achieved by wave atom transforms.[17] In this article we explore how data-driven transforms can potentially achieve sparser representation therefore higher compression than standard transforms.

3.1.1 *RPWS Time Series Compression with a 1D Wave Atom Transform*

The basic objectives of the Cassini RPWS investigation were to study radio emissions, plasma waves, thermal plasma, and dust in the vicinity of Saturn. In this study we only used RPWS time series data obtained by the Wideband Receiver (WBR) in the passband of 0.8–75 kHz. The waveform data types could produce data at prodigious rates, up to 365 kbps but more typically 60 or 183 kbps. It should be noted that the waveform data were used in analyses in two different ways. For much of the analysis the waveforms were Fourier transformed on the ground, providing much higher temporal and spectral information than that provided by the survey spectral products. Such a use of the waveforms is somewhat forgiving and could suffer from some losses in compression without loss of science utility, even though lossy compression was not employed onboard. The other use of the burst waveforms involved the use of the waveforms themselves. Two examples are examined in the following sections, and rely on distinguishing between two different types of structures that are, at first glance, quite similar. In considering compression schemes to use onboard, it was this need to deal with details of the waveforms that drove the requirement for lossless schemes. The Rice lossless coding algorithm[7] typically achieved compression ratios of $4:1$. For some RPWS time series with very high data entropy waveforms including impulsive spikes, such as those studied herein, the achieved compression was closer to $2:1$.

In this article we investigated lossy compression of RPWS time series data using the wave atom transform[13] with the constraint that the frequency and temporal structure of plasma waves are preserved with sufficient accuracy when assessed by scientists. The wave atom transform was introduced as a variant of wavelet packets and has been shown to be more advantageous in providing trade-offs between frequency resolution and time resolution. It was shown that warped oscillatory functions, such as the waveform pulses caused by dust impacts we investigated herein, have a significantly sparser expansion in wave atoms than in other fixed standard representations like wavelets and Gabor transforms. If the filters are specifically chosen, the transform can be made orthogonal.[18] Furthermore, wave atoms can be computed using Fast Fourier Transforms (FFTs), making it a good candidate for onboard processing if the mission team chooses to have a hardware chip. We include a brief description of wave atom basis construction in Appendix A. Detailed derivation and proofs can be found in reference 13.[13]

3.1.2 *ISS Image Compression using 2D Wavelets and Wave Atoms*

The Cassini ISS was the highest-resolution two-dimensional imager on the Cassini Orbiter and had been designed for investigations of the bodies and phenomena within the Saturnian planetary system. The cameras produced 12-bit data numbers (DN), stored as 16 bit words on the ground. The ISS lossy compression was a variant of the JPEG algorithm which performed discrete cosine transform (DCT) on 8×8 pixel blocks followed by quantization and Huffman encoding. Lossy compression was used for images requiring higher compression ratios and was only available to data that had been converted to 8-bit format. However, the lossy ISS data compression was shown to have introduced visible blocking artifacts and degradation of high spatial frequency components[1] even with the least amount of compression. The blocking artifacts appeared as a regular pattern of visible block boundaries. This degradation was a direct result of the coarse quantization of the coefficients and of the independent processing of the blocks. For this reason, lossy JPEG compression was used mainly during the early phase of the mission and lossless Huffman encoding was used for majority of the images.

The compression ratio of the Huffman lossless compressor varied from image to image. Image scenes with high entropy never compressed greater than $2:1$, in which case part of the data was truncated. The users would then need to interpolate the missing pixels for the spatial neighbors. This means that, where this occurred, the compression was no longer truly lossless. And often spatial down sampling (by a factor of 2 or 4) had been used to conserve bandwidth when higher compression ratio was needed. This often occurred in images that contain a lot of surface details which would have provided valuable scientific information if a more effective compressor were available.

In this article we investigate transform compression of ISS images using 2D wavelets and wave atoms. ISS observations were designed to address a multitude of Saturn-system scientific objectives and the images demonstrated large variations of spatial content and complexity. We selected a small set of the images containing Saturn, Saturn's rings, Titan (in different phases), and Iapetus (in high resolution). These images included a representative set of observations with different entropy. We evaluated the reduction in compression rates that can be achieved in lossless compression using wavelet and wave atom transforms. We also assessed the extent to which lossy compression can be tolerated without compromising the science contents of the images.

3.1.3 UVIS Spectra Compression using SVD

The UVIS sensor, a microchannel plate with a coded anode array, counts photons. The dominant noise is photon shot noise that follows a Poisson distribution. The data set we selected consists of long ($20\,\mathrm{min}$) integrations to collect as many photons as was practical given the rate of change of viewing geometry. Although in many cases spectra over a many-hour observation can be co-added to improve the signal/noise ratio (SNR), for this study we used many $20\,\mathrm{min}$ single integrations to test how well such spectra could be compressed.

We investigated Singular Value Decomposition (SVD) method for compression of UVIS spectra data. SVD is data dependent and closely related to the Karhunen–Loève transform (KLT),[18] which is the optimal transform in terms of most sparse representation for a known stationary process. SVD has been be applied to data compression[14,19–22] and, as our study shows, provides high compression ratios for a class of data types. It is especially efficient on large data sets whose data numbers can be expressed in terms of a few eigenvectors (the signal) and with a substantial contribution from noise. The technique can also be considered as a denoising method, and the lossy compression consists of returning the signal and not the noise.

For this study we examined SVD compression of Saturn spectral data where the observed photons are from scattered sunlight from the (mostly) hydrogen atmosphere. Although the spectra can depend on latitude, phase angle, and emission and incidence angles, the Saturn spectral signatures are dominated by acetylene absorption features and broad-band UV-absorbing haze. Therefore SVD provides an optimal method for energy compaction for this class of spectral data. In the context of mission operations the SVD eigenvectors will change for different targets due to variations in their absorption features. The eigenvectors would need to be computed once for each target.

3.2 Predictive Coding

The principal motivation behind predictive coding is to remove redundancy between neighboring pixels, thereby reducing the dynamic range of the data via the *predictor*. The differences between

the estimates and the actual sample values, i.e., prediction errors, are quantized (if lossy compression) and entropy coded in the compressed bit stream. The predictor design will directly affect the statistics of the prediction error and therefore the compressed bit rates.

In this article, we investigate the compression performance of JPL's Fast Lossless EXtended (FLEX) hyperspectral/multispectral image compressor on Cassini's VIMS data cubes. The FLEX compressor represents a particularly effective way of using low-complexity adaptive filtering[23,24] for predictive coding of hyperspectral images and provides an extension to the lossless algorithm[25] to perform near-lossless compression. Figure 1 shows the block diagram of the FLEX compressor.

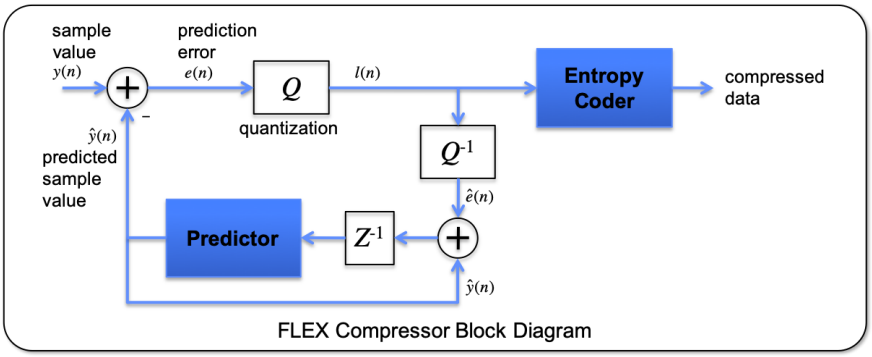


Fig 1 Block diagram of JPL's FLEX compressor.

The predictor uses an adaptive linear prediction method[25] to predict the value of each image sample based on the values of nearby samples in a small three-dimensional neighborhood. The prediction residual is quantized, and the quantizer output is mapped to an unsigned integer which is encoded using the entropy coder. The prediction mode, as well as the quantization step size, are user specified parameters which can be customized to adapt to the individual usage and needs of the data. Users can control quantizer step size by specifying an absolute error limit, so that samples can be reconstructed with a user-specified bound on reconstruction error. The compression is considered "near-lossless" when the quantization step size is smaller than data granularity or falls under the instrument intrinsic noise level.

3.2.1 VIMS Image Cube Compression using FLEX

We chose FLEX for VIMS hyperspectral data compression due to its demonstrated performance on data from more than 14 different Multi-Spectral and Hyper-Spectral imaging (MSI/HSI) instruments.[26] In July 2018, the FLEX compressor began use onboard the International Space Station as part of NASA's ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) instrument. And it has also been adopted by the Earth Surface Mineral Dust Source Investigation (EMIT) Spectrometer which is scheduled for launch in 2022.

The Cassini VIMS was composed of 2 instruments co-aligned: VIMS-V observed the visible part of the spectrum in 96 channels ($0.35\,\mu m$ to $1.04\,\mu m$) and VIMS-IR covered the infrared range ($0.88\,\mu m$ to $5.12\,\mu m$) with 256 channels allowing a small overlap around $1\,\mu m$.[3] Both instruments observed a slightly different field of view with a maximum of 64×64 pixels in the spatial dimension. Onboard the Cassini spacecraft, VIMS data were compressed using a Rice encoder[7] which achieved lossless compression ratios between 2 and 3 depending on the entropy

of the data. The native dynamic range of the instrument was 12 bits and were stored uncompressed on the Planetary Data System (PDS) image node as `.qub` files in ISIS2 format (pds-imaging.jpl.nasa.gov/volumes/vims.html).

During the 13 years of the mission, a total of 739,898 cubes were acquired representing a volume of more than 250 GB of uncompressed raw data. For this study, we selected 21 image cubes acquired during 10 Titan flybys. This data set includes observations of Titan at different spatial and spectral resolutions. We assessed data compression using FLEX lossless and near-lossless compression as compared to the onboard Rice compressor, and evaluated the impact of lossy compression by comparing the compression error with the intrinsic noise of the VIMS data.

4 Science Evaluation, Iteration, and Evolutionary Improvement

In order to reach maximum data compression, constrained by science requirements, we had to make improvements to our initial attempts based on evaluation of the products. For lossy compression there is potential for science loss. In order to find the limits where science loss can be tolerated for the RPWS data, members of the RPWS team evaluated science content of the decompressed data. They found that, after the first attempt the plasma wave signatures were retrieved with high fidelity. However, other events such as dust grains impacting the instrument and triggering a detectable response were lost in the compression. In response we applied rate allocation which adjusted the number of coefficients to be kept for each of the time series according to the signal power variations over time.After implementing this modification both the plasma wave and dust impact signatures were retrieved with high fidelity at the same compression bit rate as the first iteration.

In order to achieve the same result during a future flight mission the science team would evaluate initial results, revise the compression algorithm, and upload the new algorithm to the spacecraft. This methodology goes beyond the current paradigm for flight instrument data compression. Note our current rate allocation approach is based on minimization of mean squared error, which may not be the optimal metric to represent the scientific information distortion of the decompressed data. Our ultimate goal is to demonstrate a methodology to design compression algorithms for future outer solar system space missions able to directly incorporate scientifically motivated distortion metrics in the compression scheme. This will enable us to develop a framework to generalize the classical Shannon rate-distortion theory to optimize scientific retrieval performance of the reproduced data. We provide in section 6 an example of this kind of distortion metric for future development.

In terms of science requirements, a test of the fidelity of the retrievals using the lossy compression (as we did for the RPWS test case) would be the gold standard. For ISS, UVIS, and VIMS cases we presented a proxy method – a comparison of error introduced by the lossy compression and intrinsic noise in the data. The intrinsic noise can be estimated by the difference between the original data and the median value of the 3×3 surrounding pixels in the spatial domain. This gives us an estimate of the expected variability of the data that can be compared with the compression error. It seems reasonable to argue that if the error introduced by the lossy compression can be constrained to be some fraction of the intrinsic noise, the degradation of the retrievals would be tolerable. The threshold fraction would be determined by the science team.

For the VIMS data we presented results for the FLEX method. We are also in the process of testing the SVD method for VIMS. The comparison of SVD for UVIS and VIMS reveals some

interesting results. We found that, under SVD compression, the UVIS data can be compressed by a factor $\sim$70. For VIMS the compression factor is much lower, $\sim$10 or less depending on the science. The UVIS Saturn data compress strongly because the intrinsic SNR is relatively low, and the signal can be well represented by only three eigenvectors. The SVD compression does a good job separating signal and noise, and the compression retains only the signal to high degree. That is not the case for the VIMS cubes. For those the SNR is very high, and spectral signatures $\sim 1\,\mathrm{DN}$ can be detected in the weakest regions. It is much more difficult to get a high compression ratio under this circumstance and we are still working on it.

For the UVIS SVD analysis we used the first 400 spectra of Saturn to generate the eigenvectors, and we used those for the remaining 8000 spectra. When we tried that with the VIMS cubes we found that the data compressed well for the first cube, then deteriorated rapidly for the other cubes that span an interval of several years during the mission. This can be explained by a secular drift in the wavelengths of the channels for VIMS. The SVD method is sensitive to that. Had the VIMS team been using the SVD algorithm we speculate that they would have discovered the wavelength drift much earlier than they did.[27] We found that the SVD method can still be used if new eigenvectors are computed for each cube, or re-computed periodically during the mission.

The experiments with the VIMS data suggest hybrid approaches using two or more compression schemes, where for example, FLEX or SVD may be favored for different VIMS data cubes. To implement this flexibility on the spacecraft suggests that a facility compressor is called for, one that has a suite of different compression algorithms that can be called upon by any of the science teams depending on circumstances. A facility compressor enables other advantages. It could be supported by a team of compression experts. It could be designed for easy and rapid editing and uploading of new algorithms and compression parameters, and would alleviate science team development costs of designing their own compressors.

5 Experimental Results

In this section we report the results of applying the selected compression algorithms on Cassini data sets. To assess the impacts of lossy compression on science retrieval, we performed such an evaluation for a small part of the Cassini RPWS (Radio Plasma Wave Science) data set where the science (the detection of waves and dust particle impacts) could be easily assessed. However, for the other data sets we opted to compare lossy compression errors against intrinsic noise of the data. The rationale being that, if the compression errors are smaller than the instrumental noise it is likely that the science loss would be acceptably small.

5.1 Compression Results of RPWS Time Series

One of the main scientific objectives of RPWS observation was to detect and analyze the short-lived sporadic broadband radio emissions. The time series data from the RPWS Wideband Receiver (WBR) exhibited non-uniformly spanned information in both time and frequency dimensions. In this section we used two data sets obtained by the Cassini WBR at $75\,\mathrm{kHz}$ filter bandwidth. We applied 1D wave atom transforms independently to each of the raw time series data over $\sim 74\,\mathrm{ms}$ intervals. The transform coefficients, or a selected subset of them, are sent to an arithmetic entropy coder to produce compressed bit streams. Figure 2 shows the block diagram of the wave atom transform compression scheme used for RPWS data in this experiment.

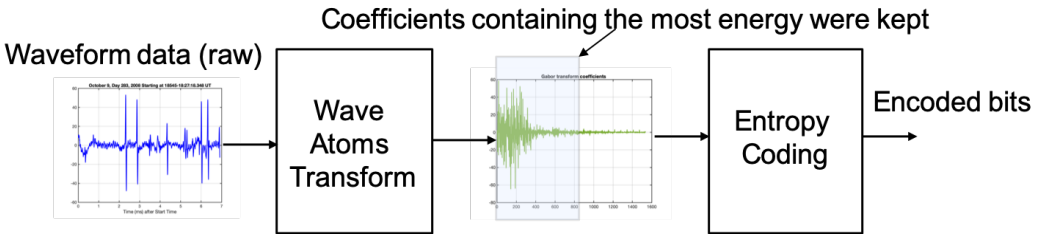


Fig 2 Block diagram of wave atom transform compression scheme for the RPWS time series data.

As shown in Figure 2, if chosen properly, the transform projects the signal into a domain where most of the energy is concentrated in fewer dimensions (i.e., sparse representation). A user can choose to keep all (lossless) or a subset of the coefficients (lossy) and send them to the entropy coder. To achieve different compression ratios, we varied the *threhsold* value which dictates the number of coefficients to be kept and encoded. The coefficients with larger magnitudes than the threshold are called *significant*. The compressed bit rates are estimated by two approaches as described below:

I Entropy coding of the *significant* coefficients only, along with the encoded significance map.

II Entropy coding of all the transform coefficients with *insignificant* coefficients set to zero.

With approach I, the significance map (a sparse binary map) needs to be encoded and transmitted as side information in order for the decompressor to perform an inverse transform accordingly.

The decompression is simply an inverse process of the blocks shown in Figure 2, i.e., inverse entropy coding followed by an inverse wave atom transform. To assess the effects of lossy compression on the scientific contents of the data, we performed calibration on both the original and the reconstructed raw waveform data to obtain the electric field waveforms. We then evaluated 1) the correlation coefficients between the original and reconstructed electric field waveforms; and 2) the ability of the compressor in preserving the time and magnitude resolutions for different types of waveform pulses.

We focused on two data sets obtained on October 9, 2008 over a $1.5\,\mathrm{h}$ period before and during an Enceladus encounter.[28] The case we examined is one of greater complexity than the majority of the RPWS data. It contains the main waveform types that the RPWS team considers to be important, e.g., a variety of plasma waves and two types of signatures from particle impacts. The average lossless compression ratio achieved by a software implementation of the Rice compression is $\sim 2:1$ for these two data sets, indicating the high entropy content in the RPWS data recorded during this time period.

Table 1 shows the compression ratios achieved by retaining different fractions of the transform coefficients and the respective correlation factors between the reconstructed electric field data and the original electric field data.

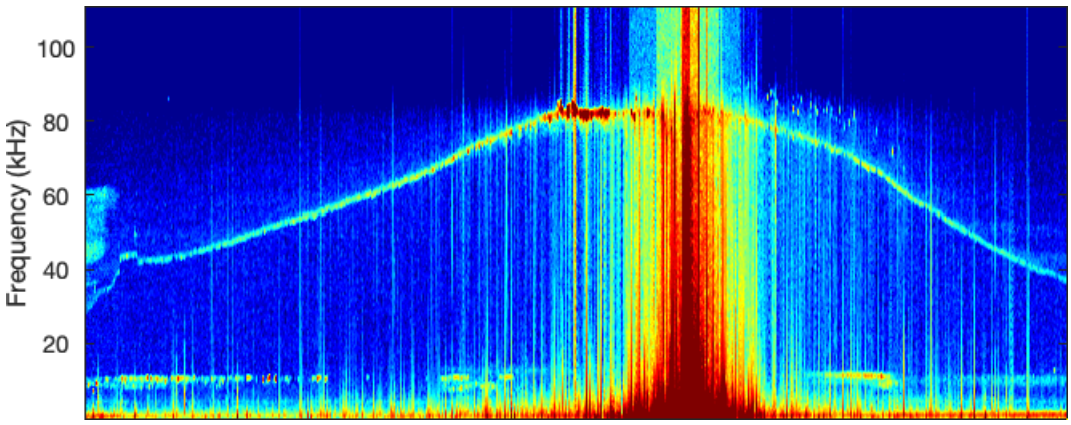
We can see that wave atom transform led to a fairly sparse representation of the RPWS time series data. The correlation coefficient between the reconstructed and the original is very high, e.g, $99.92\,\%$, even when only $15\,\%$ of the transform coefficients were used for reconstruction. The high correlation coefficients suggest that for compression ratios reported above, Fourier transform of the reconstructed waveforms will give us acceptable spectrograms, which is a major use of the

Table 1 Summary of wave atom compression results on RWPS data.

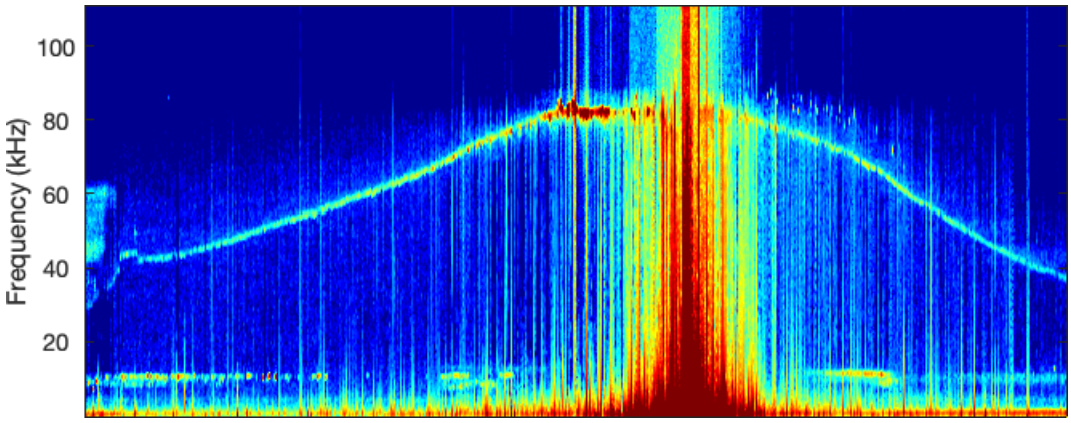
Fraction of coefficients	100 %	40 %	30 %	25 %	15 %
Approach I compression ratio	2.5 :1	4.6 :1	5.7 :1	6.6 :1	9.2 :1
Approach II compression ratio	2.5 :1	3.7 :1	4.5 :1	5.0 :1	6.8 :1
Correlation coefficient	100 %	99.99 %	99.97 %	99.96 %	99.92 %

RPWS data. For the remainder of this section, we will refer to the compression ratios of Approach I for consistency.

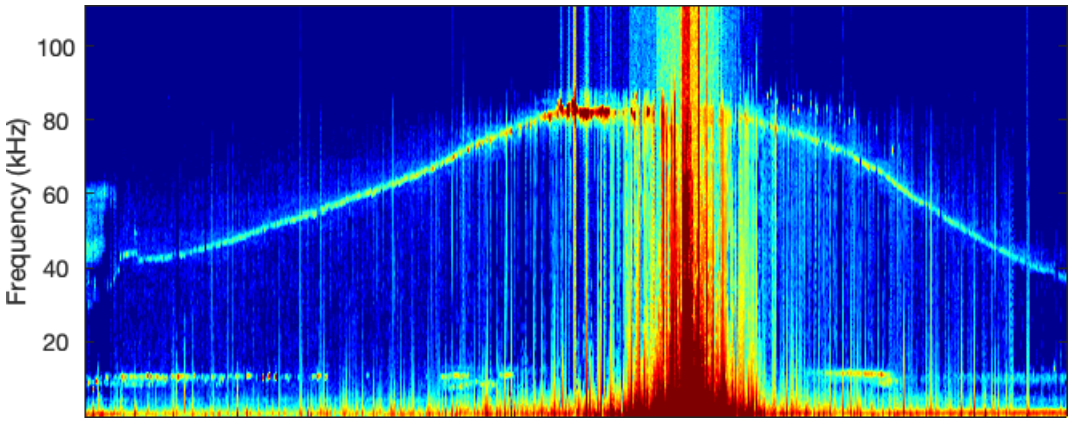
We applied Fourier transforms to all the calibrated time series contained in these two data sets (i.e., T2008283_18_75kHz and T2008283_19_75kHz) and concatenated them in time sequential order to form the frequency-time spectrogram of this period (from 18:00 to 19:30, on October 9, 2008). Figure 3 compares the spectrogram of the original waveform data (top), the spectrogram of reconstructed waveform data at compression ratio of 6.6 :1 (middle panel), and the spectrogram of reconstructed waveform data at compression ratio of 9.2 :1 (bottom panel), respectively.



(a) The original spectrogram



(b) The spectrogram of reconstructed data at compression ratio of $6.6:1$



(c) The spectrogram of reconstructed data at compression ratio of $9.2:1$

Fig 3 The frequency-time spectrograms of the original (top) and reconstructed (middle and bottom) electric field data obtained during a $1.5\,\mathrm{h}$ period event on October 9, 2008.

As shown in Figure 3, near the Enceladus closest approach RPWS detects narrowband, slowly varying emissions both near 10 kHz and also between about 40 and 80 kHz. These are the typical plasma waves that are present in Saturn's magnetosphere. The band that increases in frequency from about 40 kHz to near 80 kHz and then decreases, again, is at a characteristic frequency of the plasma dependent on the magnetic field strength and electron density. Since it is narrowband, the underlying waveforms are approximated by a sine wave of slowly varying frequency. The broadband, bursty emissions are due to impulsive events. The Fourier transform of an impulse is a broadband spectrum. As will be shown, later, there are two types of bursty emissions underlying these. First, centered near the maximum signal are multiple dust impacts. At somewhat earlier and later times, there are also electrostatic solitary waves which are unrelated to the dust. A challenge in lossy compression of these data is the preservation of our ability to distinguish between the two different types of impulses in the time domain.

Another use of the RPWS data relies directly on the waveforms themselves. For example there were a number of dust impacts present during this Enceladus flyby. The shape of the dust impact waveform was used to detect and count these events. The time and frequency resolutions need to be preserved in order to distinguish between two different types of structures that are, at first glance, quite similar. In this experiment, we investigated the ability of the wave atom compressor in reconstructing different types of dust waveforms as we push for higher compression ratios.

Figure 4 compares the original and reconstructed electric field data over a $\sim 7\,\mathrm{ms}$ interval obtained about half an hour before crossing the Enceladus dust plume. In this example, $25\,\%$ of the wave atom coefficients were kept and encoded. The resulting average compression ratio is $6.4:1$. This time series contained five well-defined bipolar ESW pulses embedded within a much lower amplitude oscillatory wave. Expanded views of the ESW waveform pulses, in the three highlighted areas of figure 4, are shown in figure 5, with the reconstructed and original waveforms on top of each other. The three types of ESW pulses, as well as the low amplitude oscillatory waves, are reproduced with high fidelity in this example.

During the closest approach to Enceladus, there was extremely intense broadband noise extending through the entire frequency range of the WBR, found to be associated with dust impacts on the spacecraft and antenna. The waveform pulses associated with the dust impacts demonstrated different scales and magnitudes than those of ESWs. Figure 6 compares the original and reconstructed electric field data over a $\sim 7\,\mathrm{ms}$ interval obtained during the closest Enceladus encounter. The three waveforms highlighted during this period are: an ESW (the light blue box), a dust impact (the light red box), and an oscillatory dust impact (the light green box). The expanded views of these three types of waveforms are shown with higher time resolution in Figure 7. The achieved compression ratio is $\sim 6.8:1$ in this example. The shapes of these three different types of waveform pulses, which are used to detect and count the dust impact events, are preserved with high fidelity.

We increased the compression ratio by reducing the fraction of retained coefficients from $25\,\%$ to $\sim 15\,\%$. Figure 8 compares the original and reconstructed electric field data over a $\sim 7\,\mathrm{ms}$ interval obtained about a half hour before crossing the Enceladus dust plume. The compression ratio achieved for this case is about $9:1$. Expanded views of the waveforms in three highlighted regions are shown in Figure 9. In comparison with figures 4 and 5, the waveform pulses of the ESWs are not impacted by increased compression. There are notable smoothing effects however, on the lower magnitude oscillatory waves when the compression ratio is increased to $9:1$ in this example.

A comparison of the original and reconstructed electric field data over a $\sim 7\,\mathrm{ms}$ interval obtained during the closest Enceladus encounter is shown in figure 10. The achieved compression

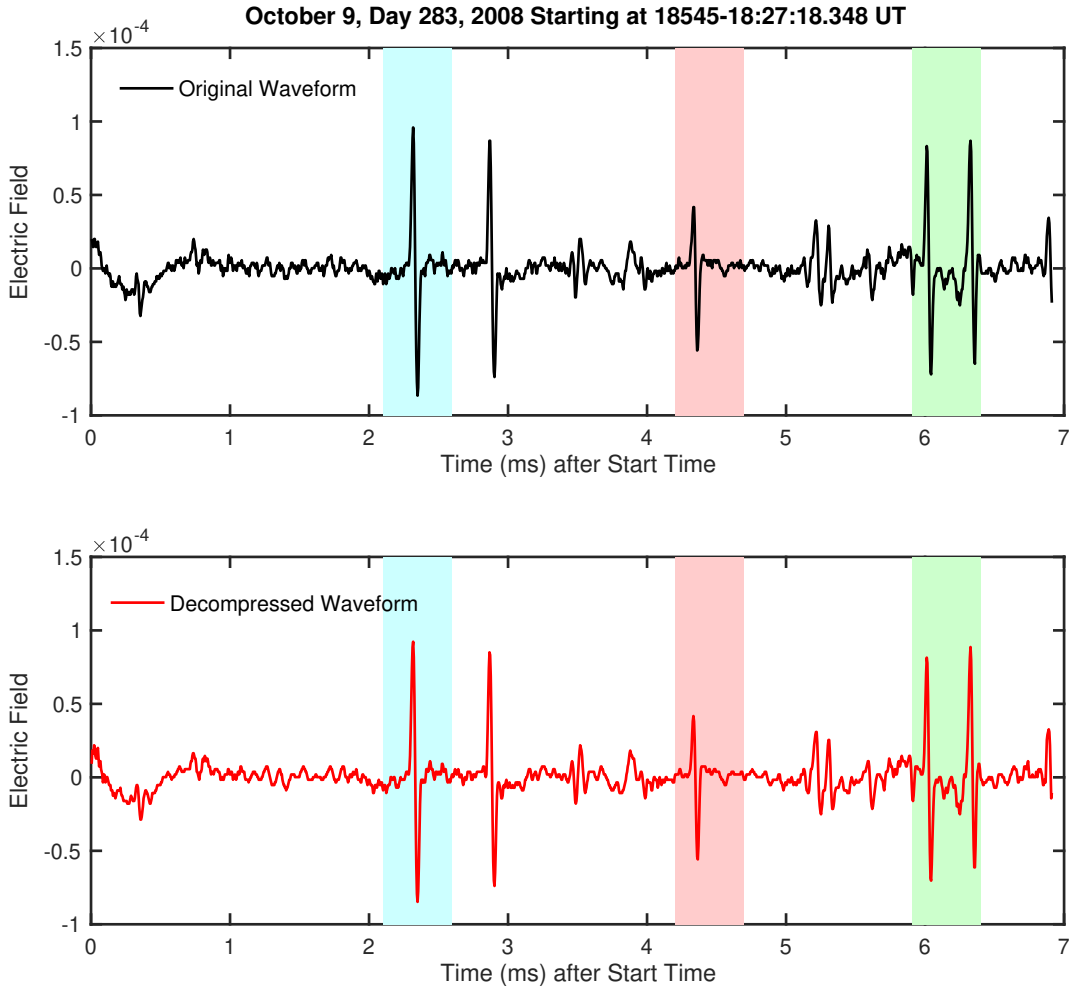


Fig 4 Original (top) and reconstructed (bottom) electric field data over a $\sim 7\,\mathrm{ms}$ interval, obtained $\sim 0.5\,\mathrm{h}$ before crossing the Enceladus dust plume. This time series contained five well-defined bipolar electrostatic solitary waves (ESW) embedded within a much lower amplitude oscillatory wave. Compression ratio is about $6.4:1$. The shaded areas contain three different ESW pulses.

ratio is $\sim 9.4:1$ in this example. Expanded views of the waveform in three highlighted areas are shown in Figure 11. The ESW, dust impact, and oscillatory dust impacts waveforms are all reconstructed with high fidelity, albeit with some smoothing effects on the lower amplitude waves.

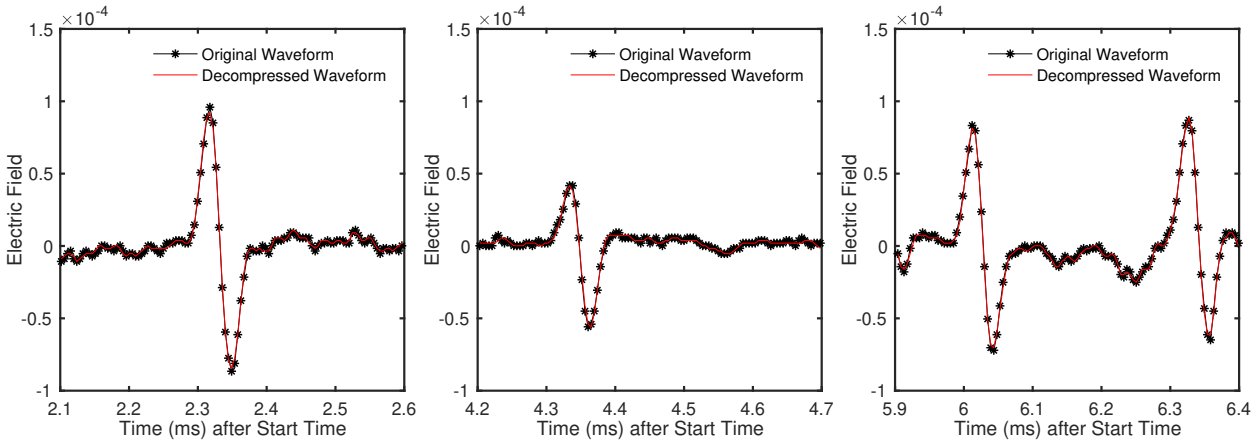


Fig 5 Expanded view of original and reconstructed ESW pulses captured $\sim 0.5\,\mathrm{h}$ before crossing the Enceladus dust plume. Compression Ratio is about $6.4:1$ in this example.

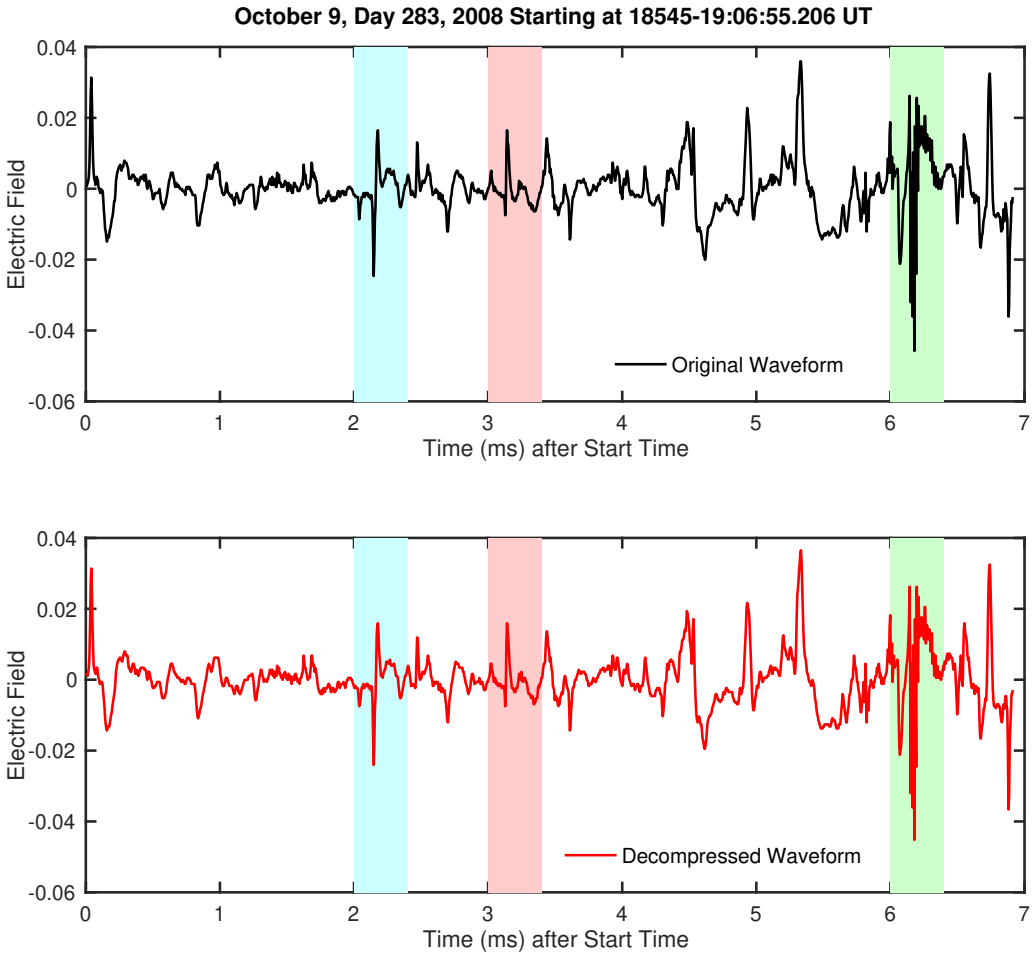
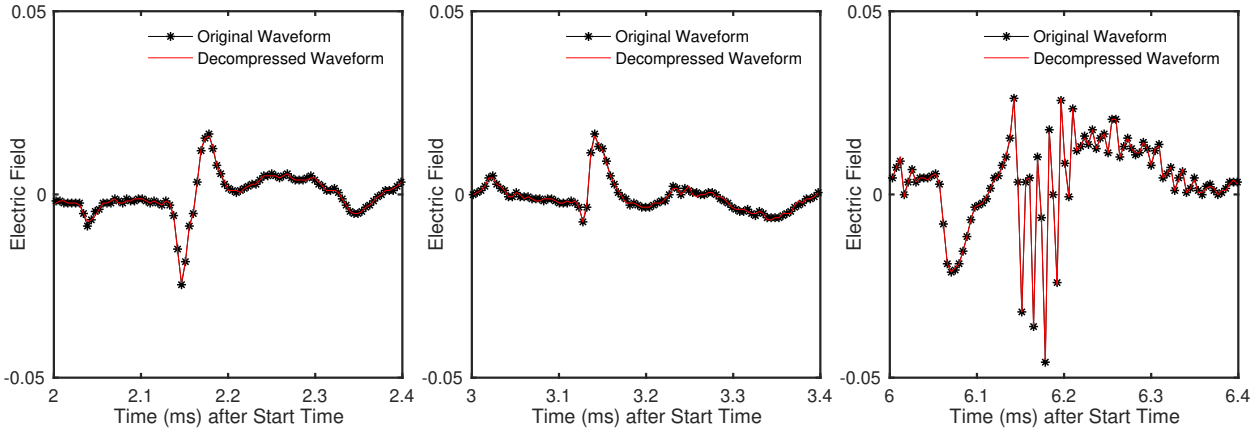


Fig 6 Original and reconstructed time series data during the closest Enceladus encounter. The highlighted areas contain three types of waveforms: an ESW pulse (the light blue box), a dust impact (the light red box), and an oscillatory dust impact (the light green box). Compression Ratio is $\sim 6.8:1$ in this example.



(a) ESW pulse (b) Dust impact (c) Oscillatory dust impact

Fig 7 Expanded views of three types of waveforms (original and reconstructed) captured during Enceladus encounter: (a) An ESW pulse showing the bipolar nature of nearly symmetric negative and positive peak; (b) Dust impact signature consistent with dust impacting the spacecraft; and (c) Oscillatory dust impact signature consistent with ringing at the plasma frequency. Compression ratio is $\sim 6.8:1$ in this example.

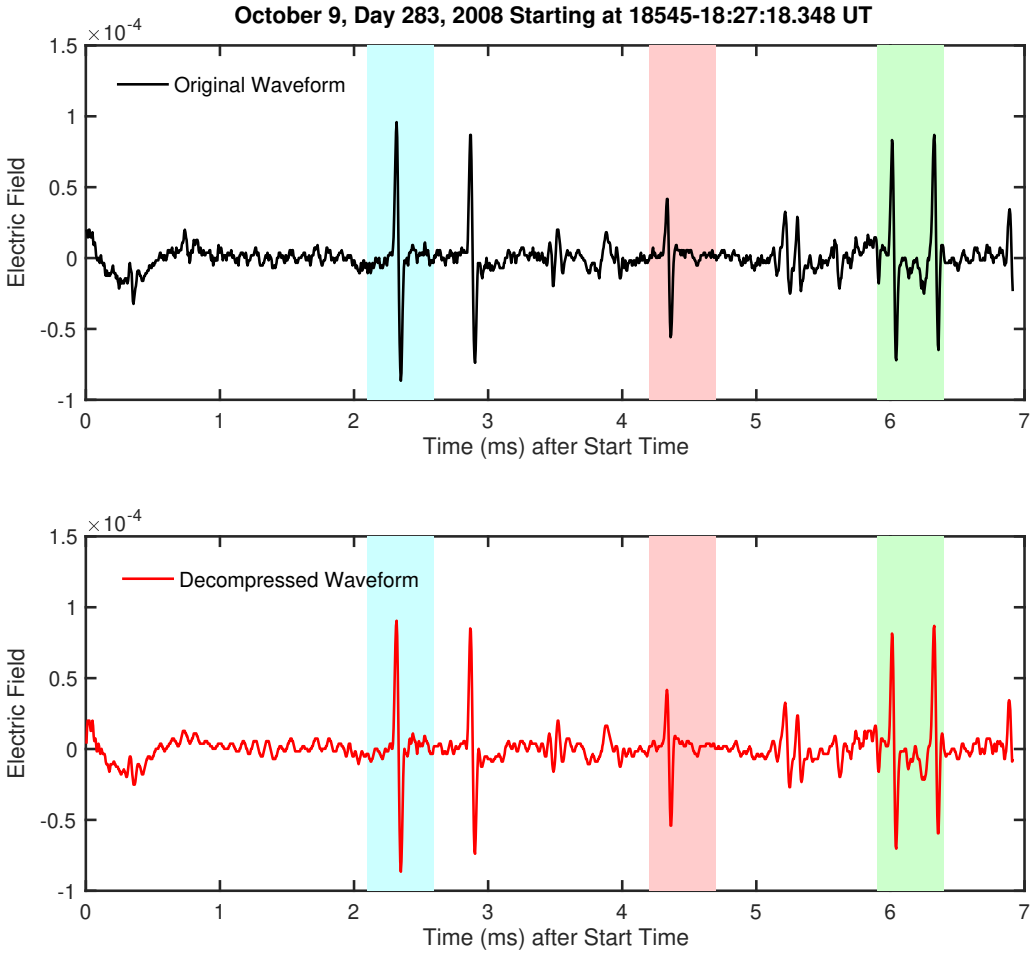


Fig 8 Original (top) and reconstructed (bottom) electric field data over a $\sim 7\,\mathrm{ms}$ interval, obtained $\sim 0.5\,\mathrm{h}$ before crossing the Enceladus dust plume. The highlighted areas contain three different types of electrostatic solitary waves. We kept $\sim 15\,\%$ of coefficients and the compression ratio is $\sim 9:1$ in this case.

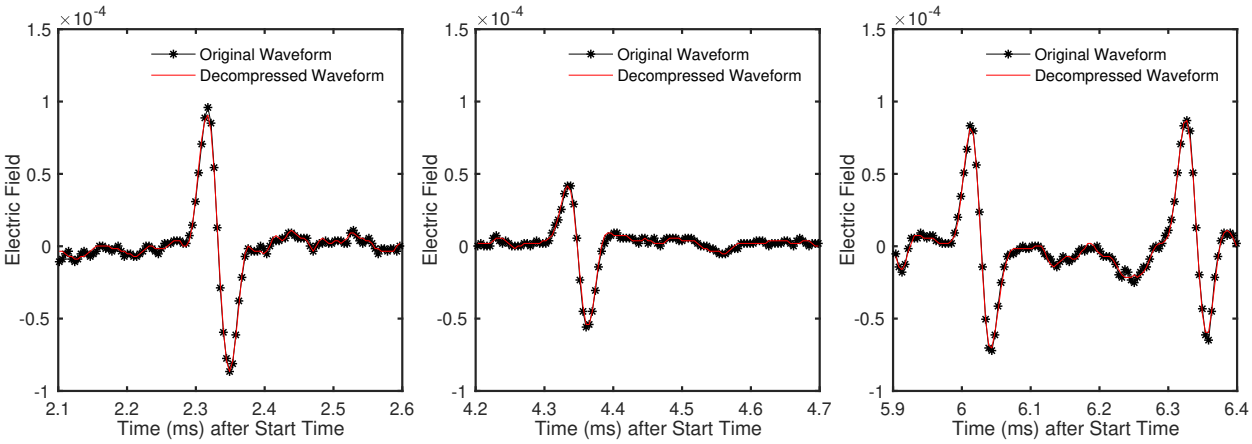


Fig 9 Expanded view of original and reconstructed waveform data from the highlighted areas shown in Figure 8.

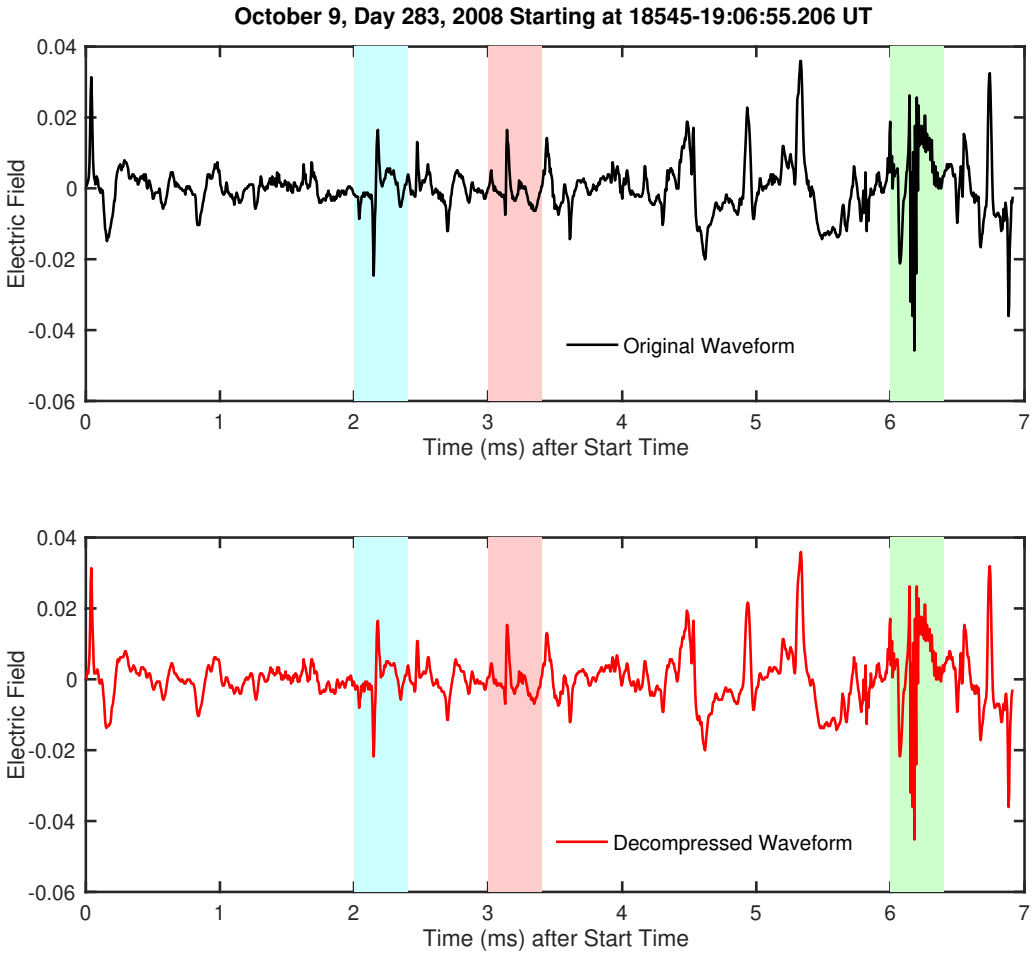
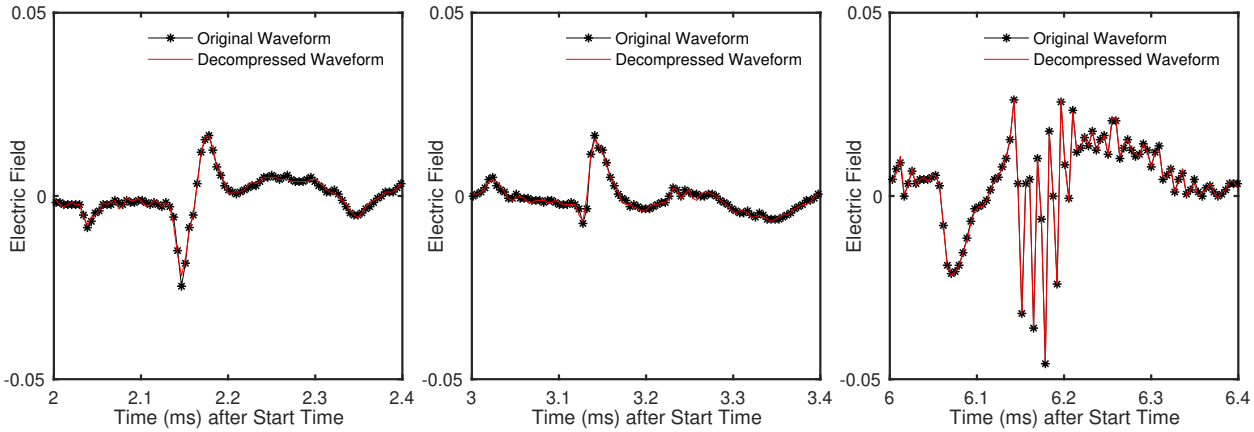


Fig 10 Original and Reconstructed Time Series data during the closest Enceladus encounter at compression ratio of $\sim 9.4:1$.



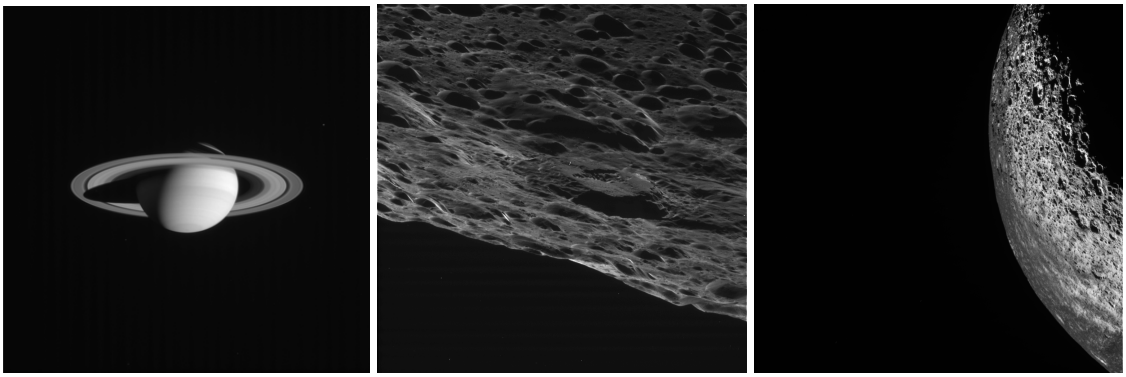
(a) ESW pulse (b) Dust impact (c) Oscillatory dust impact

Fig 11 Expanded view of original and reconstructed waveforms for the three highlighted areas in Figure 10. Compression ratio is $\sim 9.4:1$.

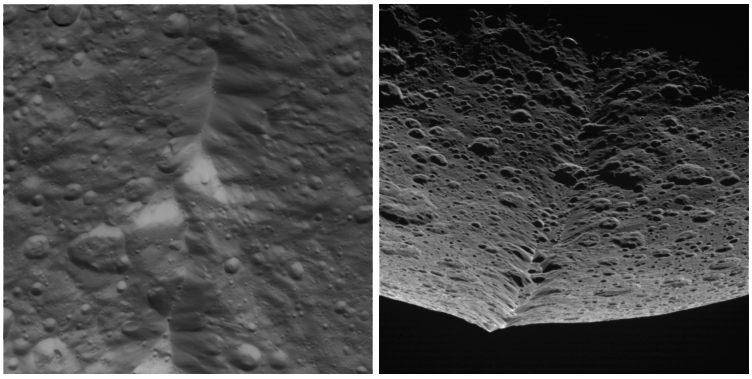
5.2 Compression Results of Cassini ISS Images

In this section we report on ISS compression experiments, with sample images including objects such as Saturn, Saturn's rings, Titan, and Lapetus. We applied threshold based quantization to either wave atom or wavelet basis transform coefficients,followed by arithmetic entropy coding to estimate achievable compression bit rates. For the wavelet transform we used Daubechies' biorthogonal symmetric filters (tap length of 5 for low-pass filter and 3 for highpass filter).[15] For the wave atom basis we used the Villemoes's construction described in Appendix A with implementation provided in the wave atom Matlab toolbox.[29] We assessed both lossless and lossy compression performance on the selected ISS images. The visual qualities, as well as objective fidelity measures, are evaluated at different compression ratios.

For the lossless compression experiment, we selected five ISS images with varying degrees of entropy content (Figure 12). These images were losslessly compressed onboard using Huffman coding and received by the ground station without any packet loss, thus providing an appropriate reference for our lossless experiment. Table 2 compared the lossless compression ratios (CR) achieved by the original onboard compressor and our proposed transform based method. The original compression ratios were extracted from the image header. On average, the wave atom transform based method achieved a factor of $\sim$1.7 improvement in data volume reduction compared to the original Huffman algorithm.



(a) N1454732522_1 (b) N1568121240_2 (c) N1483245127_1



(d) W1568127472_2 (e) W1568124782_2

Fig 12 Selected ISS images for lossless compression experiments. These images have been losslessly compressed onboard and transmitted to the ground without any packet loss.

Table 2 Lossless Compression Results on selected ISS images.

Image ID	NS × NL	Format	Original CR	Wave Atom CR	Wavelet CR
N1454732522_1	1024 × 1024	8 bit	4.4:1	11.1:1	8.1:1
N1568121240_2	1024 × 1024	16 bit	3.1:1	4.7:1	3.4:1
N1483245127_1	1024 × 1024	16 bit	3.5:1	4.1:1	4.1:1
W1568127472_2	1024 × 1024	16 bit	2.7:1	4.6:1	3.1:1
W1568124782_2	1024 × 1024	8 bit	2.7:1	4.7:1	3.3:1
Average CR			3.3:1	5.8:1	4.4:1

The ISS onboard lossless compression algorithm worked to ensure no less than a compression ratio of $2:1$. If the entropy was sufficiently high to cause the number of bits to exceed the allocation for a line pair, i.e., $1024{\times}2{\times}12$ bits, the algorithm would put 0 in all the remaining pixels of the second line of the pair. When decoding, the missing pixels values are determined by interpolating neighboring values. Effective lossless compression for images with high entropy content enables more data with higher spatial resolution to be transmitted without information loss.

For lossy compression experiments, we used 21 ISS images containing Saturn, Saturn's rings, and Titan (in different phases). We evaluated the visual qualities of the reconstructed images and compared the compression error with the intrinsic noise in the original image. The intrinsic noise was estimated as the difference between the original image and a 3×3 median filtered image. A comparison of reconstructed images from wavelet and wave atom compression for a Saturn's ring image ($N1544052794\_1$) at compression ratio of $\sim 10:1$ is shown in figure 13.

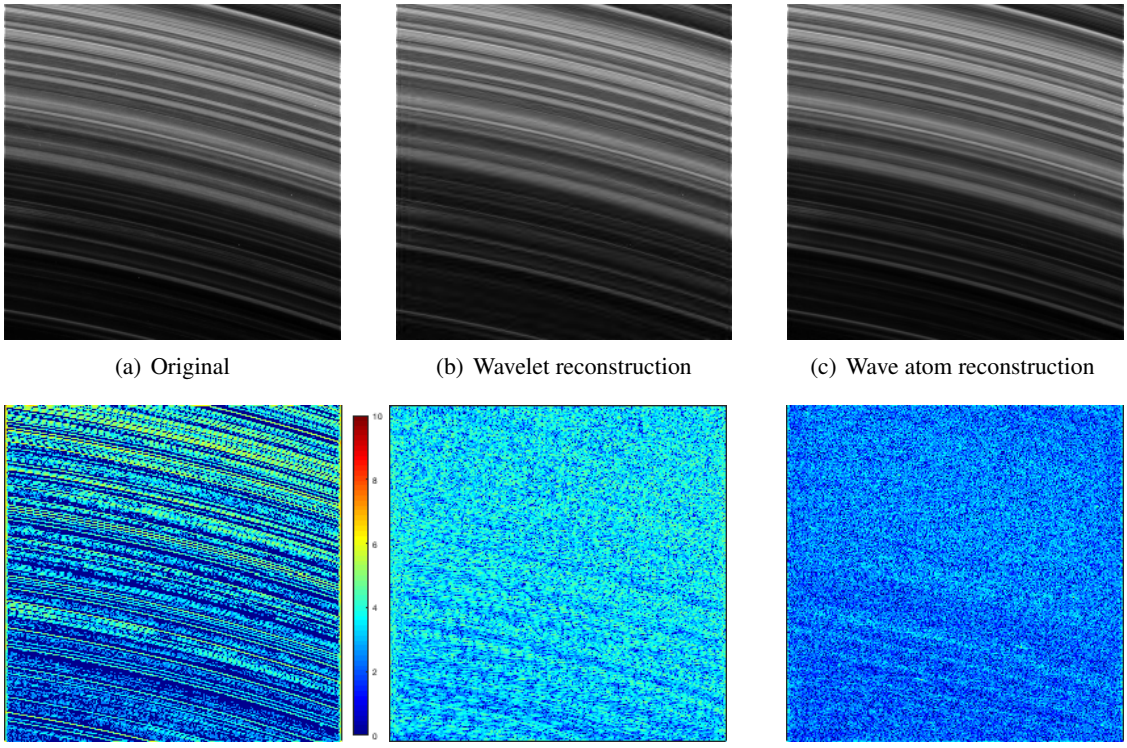
Shown in the top row (from left to right) are the original Saturn's rings image, the wavelet reconstructed image, and the wave atom reconstructed image, respectively. The bottom row shows, in logarithmic scale, the estimated intrinsic noise power and the compression noise power from wavelet and wave atom reconstructions. The visual quality of reconstructed image from wave atom transform method is almost on par with the wavelet transform method in this example. However, the wavelet method caused more degradation in high spatial frequency (smoothing of the ring patterns). This is evident in Figure 13(e) where we see higher energy tracing the rings in the image of wavelet reconstruction error. The compression error of the wave atom reconstructed image is significantly lower than the wavelet reconstructed image, and is well below the intrinsic noise level of the original image.

The rate distortion performances of wavelet and wave atom transform compression methods at various compression ratios for this Saturn's rings image is shown in Figure 14. We used the Peak Signal-to-Noise Ratio (PSNR) as a qualitative distortion metric (generally accepted as a measure of objective quality evaluation in 2D image compression) defined as:

$$\text{Peak Signal-to-Noise Ratio (PSNR)} = 20\log_{10}\left(\frac{2^b-1}{\sqrt{\text{MSE}}}\right) \quad (\text{dB}), \tag{1}$$

where b is the number of bits in the original image, and MSE is the mean-squared error between the original and reconstructed images.

In this example, the wave atom transform method demonstrates significant advantage over the wavelet transform in terms of energy compaction. This is supported by the increasing gap between the rate distortion curves towards higher compression ratios (with a smaller number of coefficients



(a) Original (b) Wavelet reconstruction (c) Wave atom reconstruction

(d) Intrinsic noise (e) Wavelet compression error (f) Wave atom compression error

Fig 13 Compression results for a Saturn's rings image ($N1544052794\_1$) at compression ratio of $\sim 10:1$. (a) the original image; (b) reconstructed image at compression ratio of $9.87:1$ with wavelet transform compression; (c) reconstructed image at compression ratio of $9.64:1$ with wave atom transform compression; (d), (e), and (f) are the estimated intrinsic noise, compression error in wavelet reconstructed image, compression error in wave atom reconstructed image, respectively. All the error images shown are in logarithmic scale which were calculated from the difference image as $20 \log_{10} |I - \hat{I}|$, where I and $\hat{I}$ refer to the original and reconctructed images, respectively.

are used for reconstruction). For example at a compression ratio of $\sim 10:1$, the wave atom reconstruction achieved a PSNR of $46.66\,\mathrm{dB}$, which is $\sim 10\,\mathrm{dB}$ higher than the wavelet reconstruction. This is consistent with the lower error in the reconstructed image using wave atoms as shown in Figure 13. Another example is shown in figure 15, of the reconstructed images from partial wavelet and partial wave atom coefficients at compression ratio of $\sim 10:1$ for a WAC Saturn image. Similar to the Saturn's rings image example, wave atom reconstruction demonstrated visibly lower error than wavelet reconstruction.

The rate distortion performance of the wave atom and wavelet methods at different compression ratios for this WAC Saturn image are compared in figure 16. The two transforms demonstrated comparable performance in high fidelity regions (lower compression ratios). However, for this image the wave atom transform consistently outperformed wavelet transform towards higher compression ratios, e.g., $> \sim 10:1$ while achieving a target PSNR. For example, the wave atom transform achieves a PSNR of $55\,\mathrm{dB}$ with a factor of 2 higher compression than the wavelet transform.

Table 3 provided a summary of lossy compression results on 21 ISS images collected during 10 Saturn flybys. PSNR (in dB) of reconstructed images at compression ratios of $5:1$ and $10:1$

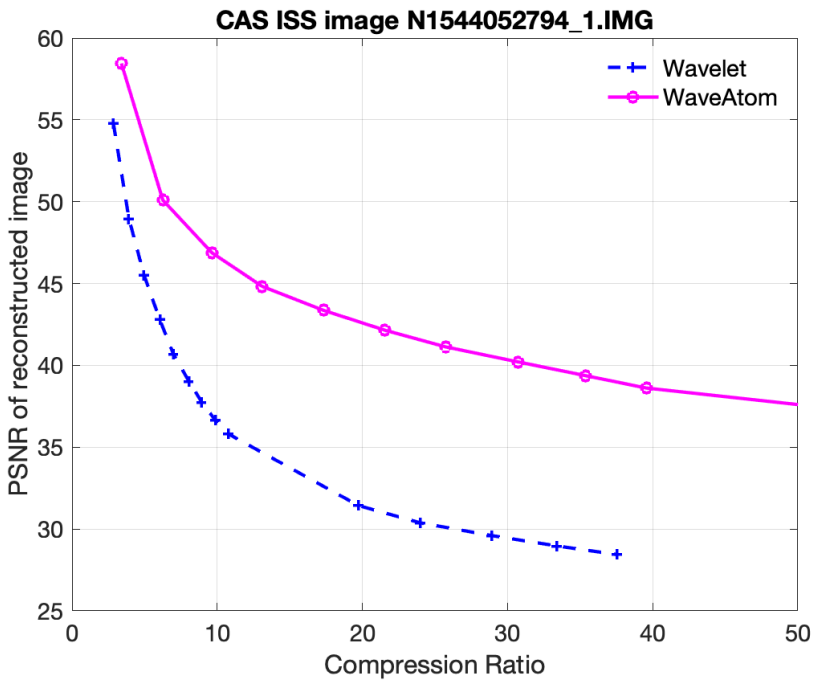
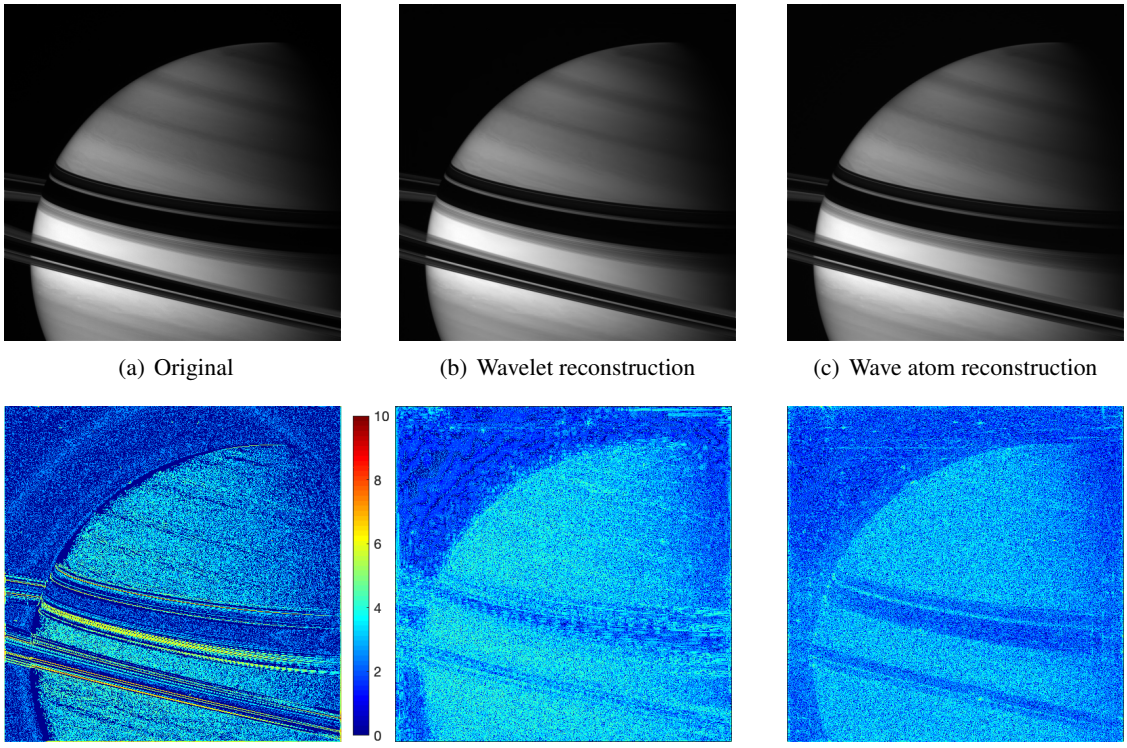


Fig 14 PSNR as a function of compression ratio for a NAC Saturn's rings image ($N1544052795\_1$) using wave atom (magenta) and wavelet (blue) based compression methods.

are shown for both wave atom transform and wavelet transform based methods. Higher PSNR values are highlighted in bold for each image. The wave atom transform based method provides a significant advantage in the compression of images with oscillatory patterns, such as Saturn and Saturn's rings. For images containing mostly smooth spatial contents, such as Titan in different phases, wavelets achieved better performance. This again motivates the concept of having a suite of compression algorithms that can be called upon to adapt to different scenarios and observations of the spacecraft.



(a) Original (b) Wavelet reconstruction (c) Wave atom reconstruction

(d) Intrinsic noise (e) Wavelet compression error (f) Wave atom compression error

Fig 15 Compression results on an ISS WAC image at compression ratio of $\sim 10:1$. (a) the original image; (b) reconstructed image at compression ratio of $10.5:1$ with wavelet transform compression; (c) reconstructed image at compression ratio of $10.8:1$ with wave atom transform compression; (d), (e), and (f) are the estimated intrinsic noise, compression error in wavelet reconstructed image, compression error in wave atom reconstructed image, respectively. All the error images shown are in logarithmic scale which were calculated from the difference image as $20 \log_{10} |I - \hat{I}|$.

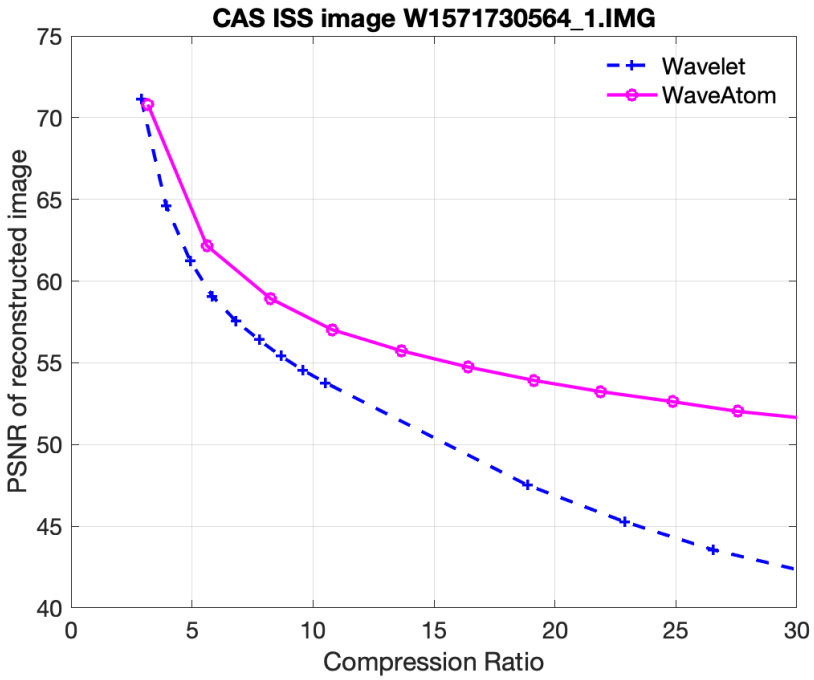


Fig 16 PSNR performance as a function of compression ratio for wave atom (magenta) and wavelet (blue) based methods on a WAC Saturn image ($W1571730564\_1$). Wave atom transform consistently outperforms wavelets, especially towards higher compression ratios.

Table 3 Lossy Compression Results on the selected ISS images.

Image ID	**NS × NL**	**Format**	**Target**	**Wave atom 5 :1**	**Wavelet 5 :1**	**Wave atom 10 :1**	**Wavelet 10 :1**
N1543939666_1	256 × 256	16 bit	Saturn-Rings	**53.05**	47.30	**46.33**	37.99
N1543939677_1	256 × 256	16 bit	Saturn-Rings	**53.40**	47.88	**46.93**	38.66
N1543939688_1	256 × 256	16 bit	Saturn-Rings	**53.92**	47.80	**47.35**	38.50
N1544049726_1	256 × 256	16 bit	Saturn-Rings	**68.17**	49.47	**60.28**	44.47
N1544049737_1	256 × 256	16 bit	Saturn-Rings	**68.06**	49.25	**60.08**	44.28
N1544049748_1	256 × 256	16 bit	Saturn-Rings	**67.74**	49.47	**59.96**	44.51
N1544052794_1	256 × 256	16 bit	Saturn-Rings	**53.77**	45.41	**46.66**	36.53
N1544052805_1	256 × 256	16 bit	Saturn-Rings	**54.01**	45.56	**46.85**	36.68
N1544052816_1	256 × 256	16 bit	Saturn-Rings	**53.33**	45.46	**46.20**	36.55
N1547134568_1	256 × 256	16 bit	Titan	45.64	**52.64**	41.71	**47.75**
N1551888649_1	256 × 256	16 bit	Titan	53.20	**60.68**	48.86	**54.89**
N1552431511_1	256 × 256	16 bit	Titan	37.10	**42.75**	33.81	**38.52**
N1577093981_1	256 × 256	16 bit	Titan	36.29	**41.40**	33.06	**37.33**
N1579532848_1	256 × 256	16 bit	Titan	36.31	**41.78**	33.04	**37.65**
W1497405997_2	256 × 256	16 bit	Software test	62.97	**67.33**	54.39	**57.01**
W1571730564_1	512 × 512	16 bit	Saturn	**64.34**	61.06	**57.62**	54.21
W1571730681_1	512 × 512	16 bit	Saturn	**61.39**	58.35	**54.60**	51.35
W1571730626_1	512 × 512	16 bit	Saturn	**65.56**	61.88	**58.70**	54.97
W1571730736_1	512 × 512	16 bit	Saturn	**61.11**	57.97	**54.23**	50.93
W1571730846_1	1024 × 1024	16 bit	Saturn	61.79	**62.41**	**56.64**	56.51
W1571730974_1	1024 × 1024	16 bit	Saturn	**60.42**	58.55	**55.00**	52.93
Average PSNR				**55.78 dB**	52.11 dB	**49.63 dB**	45.34 dB

5.3 Compression Results of UVIS Spectra Data Using SVD

For this study we examined more than 8000 ultraviolet spectra of Saturn obtained by the Cassini UVIS instrument.[4] We selected spectra of Saturn in the spectral range 168–185 nm where the observed photons are from sunlight scattered by the (mostly) hydrogen atmosphere. The spectra are dominated by acetylene absorption features and broad-band UV-absorbing haze. Although spectral signatures of other hydrocarbons are seen in the UVIS occultation spectra[30] absorption by other trace gases have so far not been detected from this data set.

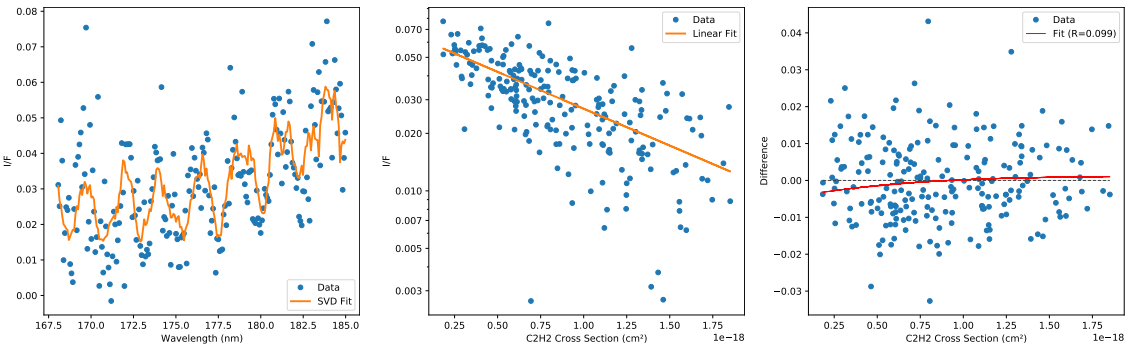
In our study the SVD operated on calibrated data which largely removes detector flat field and includes a division by the solar spectrum to obtain the relevant science quantity I/F where I is the reflected intensity and πF is the solar flux incident on Saturn's atmosphere. A compressor operating on the spacecraft could perform these operations but typically it would operate on the raw data. Our SVD compression result is insensitive to these operations because the patterns being removed are fixed, and would be present, and removable, as part of the eigenvector decomposition of the raw data. We verified this with an SVD operating on the data prior to division by the solar spectrum.

Our data sample consist of some 8902 spectra (60 spatial samples along the slit times n time steps) for each of 8 observations at a variety of locations and times. The number of time steps is different for each observation, and we discarded pixels that were too close to the limb or terminator, or had ring interference or ring shadow interference.

We can anticipate that spectra will depend on latitude (due to latitudinal gradients in acetylene abundance and UV-absorbing aerosols), phase angle, and emission and incidence angles (from radiative transfer considerations). A test of the method is therefore to derive eigenvectors from a small subset of the data, apply them to the larger data set, and measure performance. The samples span a wide range of the parameters expected to influence the measurements and therefore should constitute a robust test of the method.

One spectrum for sample 500 is shown in Fig. 17(a). As Fig. 17 demonstrates, The data show a signature of acetylene absorption and follow a linear relationship in $\log(I/F)$ as a function of the acetylene absorption coefficient. Therefore, we can take advantage of this relationship to reduce the uncertainty in the retrieval of acetylene and also of the aerosol absorption by selecting models which best reproduce the linear fit. To determine if our three-component SVD representation can be used for retrievals it suffices to show that the linear fit to the SVD representation is close to the linear fit determined from the data, within an acceptable error limit. For that purpose, we mapped the differences (data value - fit) as shown in Fig. 17(c) and computed the standard deviation. We also computed the standard deviation of the values from the SVD representation fitted to a straight line in $\log(I/F)$. The ratio of these standard deviations we label $R = \sigma_S/\sigma_D$ where σ_S is the standard deviation for the linear fit using the three-vector SVD representation, and σ_D is computed from the data shown in Fig. 17(c). The metric R is a measure of the RMS error introduced by the SVD compression relative to the RMS statistical variance in the data. We expect that retrievals of parameters from the SVD representation will be sufficiently close to retrievals from the uncompressed data provided $R < 1$, to be determined by the science team. For the case shown in Fig. 17(c), σ_D = 0.011, a typical value, and $R = 0.099$.

Fig. 18 plots ratio R of the standard deviation for the three-component SVD fit to the standard deviation of the Saturn data relative to a linear fit (Fig. 17(b)). Each sample is a full spectrum as shown in Fig. 17(a).



(a) Spectrum (b) C2H2 absorption (c) Difference

Fig 17 Sample 500 (latitude $-59.8°$, emission angle $10.7°$, incidence angle $71.0°$). Fig. 17(a) shows the spectrum for sample 500 (latitude $-59.8°$, emission angle $10.7°$, incidence angle $71.0°$). Data are plotted as + symbols, the curve is the spectrum formed from the first three SVD eigenvectors. In Fig. 17(b) the data are plotted as a function of acetylene absorption cross section (cm^2).[31] A best-fit straight line to $\log(I/F)$ as a function of acetylene cross section is also shown. In Fig. 17(c) the difference in I/F between the data and the linear fit are shown, along with a curve calculated from a linear fit, as in Fig. 17(b), but using the SVD 3-eigenvector amplitudes instead of the data (solid red curve). The dash line shows the zero-difference location. The ratio ($\sigma_S/\sigma_D = R$) is indicated, where σ_S is the standard deviation for the linear fit using the three-vector SVD representation relative to the best-fit straight line computed from data, and σ_D is computed from the data shown in Fig. 17(c).

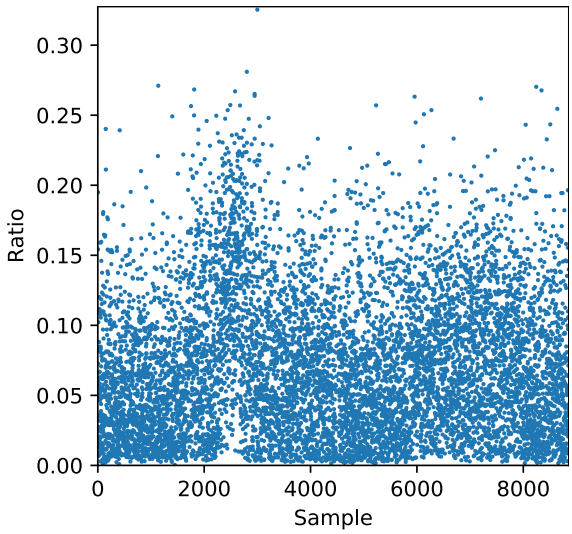


Fig 18 The ratio R of the standard deviation for the three-component SVD fit to the standard deviation of the Saturn data relative to a linear fit (Fig. 17c.)

The cumulative fraction of samples as a function of R is shown in Fig. 19. The $99\,\%$ line corresponds to $\sigma_S/\sigma_D = 0.18$. The other $1\,\%$ have ratios less than 0.32. Those numbers support our conclusion that the SVD representation of the data using only three eigenvectors would be sufficient for retrievals. A further implication is that the additional 215 eigenvectors account for the noise. The SVD procedure can therefore be considered a denoising algorithm. The resulting compression factor is $218/3 = 70.3$.

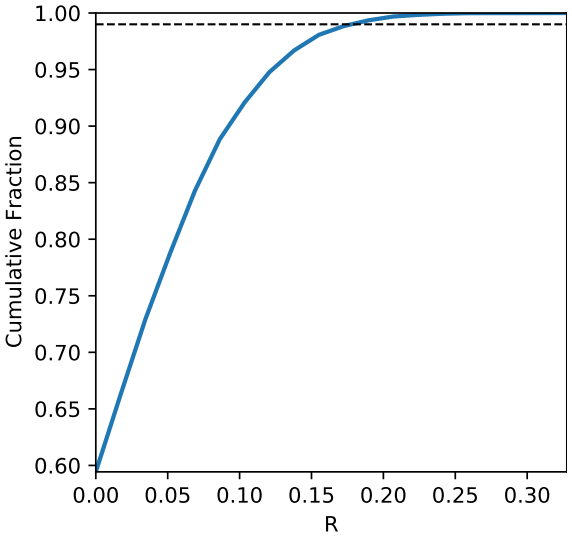


Fig 19 The cumulative fraction of cases with σ_S/σ_D (R) less than the value on the abscissa. The 99 % cumulative fraction is shown as a dashed line.

The SVD algorithm can be computationally expensive. An implementation strategy would therefore put the computation of the eigenvectors on the ground rather than on the spacecraft where the computational resources are limited. In the example described above the first 400 spectra of the sample set were used to construct the 218 eigenvectors. In the context of mission operations these 400 spectra would be transmitted to the ground using lossless compression or no compression. The eigenvector amplitudes were calculated for each spectrum, including the $8000+$ spectra not used in the calculation of the eigenvectors by taking the dot product of each eigenvector with the data vector. In the context of flight operations, the three eigenvectors would be transmitted to the spacecraft and stored in memory. Dot products for three eigenvectors would be computed on the spacecraft. That operation is not expected to drive computational resources. The three amplitudes for each spectrum would then be transmitted to ground instead of 218 data values. To ensure data integrity, to estimate how many eigenvectors would suffice, and to evaluate results in real time science teams would need to be able to download uncompressed or losslessly compressed data periodically and be able to upload more or different eigenvectors periodically. Different targets would require a different set of eigenvectors and these would need to be stored in memory and applied according to which target is being observed.

5.4 Compression Results of VIMS data using FLEX

Titan, the largest moon of Saturn, was present in at least $33{,}500$ resolved VIMS cubes.[32] In this study we used a small fraction of this subset (21 cubes), acquired during 10 Titan flybys between 2008 and 2015. We assessed the compression performance of JPL's FLEX compressor[26] on VIMS data with four levels of increasing compression ($e = 0, 1, 2$ and 3), where e is the maximum absolute error parameter used by FLEX compressor to bound distortion of compression. We will use e_k as an abbreviation to $e = k$. Table 4 contains the list of the cubes analyzed, some acquisition parameters, the size of the uncompressed raw data, original Rice compression ratios, and compression ratios achieved by FLEX compressor.

In each case, the lossless compression ratios with FLEX ($e = 0$) are higher than the original Rice compression ratios. On average, FLEX is able to provide almost a factor of two improvement in data rate reduction (i.e., $5.4:1$ vs. $2.8:1$) in lossless compression. To assess to what extent lossy compression can be applied without compromising valuable information, we reformatted the decompressed raw data to ISIS2 format to be calibrated with the standard VIMS calibration pipeline (RC19) to derive the radiance factor I/F.[27]

Table 4 Summary of compression results on a subset of VIMS cubes.

Image ID	Date	Flyby	NS × NL	Distance	Pixel scale	Size	Original CR	FLEX e_0	FLEX e_1	FLEX e_2	FLEX e_3
1590648776_1	2008-05-28	T44	28 × 14	43,971 km	15.6 km	0.5 MB	2.9:1	3.6:1	5.3:1	6.9:1	7.8:1
1732875312_1	2012-11-29	T88	64 × 64	6546 km	3.6 km	4.4 MB	3.3:1	6.5:1	15.4:1	32.3:1	47.2:1
1732881586_1	2012-11-29	T88	64 × 16	40,725 km	14.5 km	1.2 MB	2.8:1	5.3:1	9.9:1	16.5:1	20.9:1
1764547231_1	2013-11-30	T96	36 × 64	25,241 km	13.7 km	2.6 MB	2.6:1	5.1:1	9.4:1	15.0:1	19.8:1
1764549705_1	2013-11-30	T96	60 × 62	15,102 km	7.9 km	4.0 MB	3.4:1	6.6:1	15.4:1	31.4:1	45.6:1
1767272325_1	2014-01-01	T97	40 × 30	190,899 km	67.2 km	1.4 MB	2.6:1	4.8:1	7.5:1	11.7:1	14.9:1
1767274981_1	2014-01-01	T97	20 × 20	177,470 km	62.3 km	0.6 MB	2.5:1	3.9:1	5.4:1	7.4:1	8.4:1
1767299935_1	2014-01-01	T97	34 × 34	42,465 km	21.2 km	1.4 MB	3.2:1	5.7:1	11.0:1	19.0:1	24.2:1
1767302582_1	2014-01-01	T97	64 × 64	26,192 km	13.6 km	4.4 MB	3.1:1	6.2:1	13.9:1	26.8:1	38.8:1
1767303982_1	2014-01-01	T97	64 × 64	18,486 km	9.9 km	4.4 MB	3.1:1	6.3:1	14.1:1	28.0:1	41.4:1
1784582201_1	2014-07-20	T103	38 × 38	184,727 km	65.5 km	1.7 MB	2.2:1	4.2:1	6.1:1	8.8:1	10.8:1
1787271237_1	2014-08-20	T104	64 × 64	168,828 km	84.5 km	4.4 MB	2.9:1	5.7:1	10.8:1	19.0:1	27.1:1
1787307197_1	2014-08-21	T104	64 × 64	19,716 km	10.7 km	4.4 MB	2.8:1	5.5:1	11.0:1	18.7:1	26.0:1
1787314297_1	2014-08-21	T104	64 × 64	59,981 km	32.6 km	4.4 MB	2.1:1	4.0:1	6.0:1	8.2:1	10.2:1
1790050280_1	2014-09-22	T105	64 × 64	38,599 km	19.9 km	4.4 MB	3.1:1	6.3:1	14.3:1	28.2:1	41.1:1
1790050948_1	2014-09-22	T105	64 × 64	34,674 km	18.1 km	4.4 MB	3.1:1	6.3:1	14.0:1	26.9:1	37.8:1
1790056808_1	2014-09-22	T105	64 × 64	3005 km	2.3 km	4.4 MB	2.9:1	6.0:1	12.8:1	23.6:1	33.9:1
1799682855_1	2015-01-11	T108	64 × 40	85,841 km	44.5 km	2.8 MB	2.5:1	4.5:1	7.5:1	10.9:1	13.9:1
1799684730_1	2015-01-11	T108	60 × 46	74,460 km	39.6 km	3.0 MB	2.4:1	4.6:1	7.6:1	11.4:1	15.0:1
1805207246_1	2015-03-16	T110	64 × 64	14,403 km	8.3 km	4.4 MB	2.8:1	5.6:1	11.5:1	20.1:1	28.7:1
1809727868_1	2015-05-07	T111	64 × 64	28,345 km	14.9 km	4.4 MB	3.1:1	6.2:1	13.9:1	27.1:1	39.4:1
Average CR							2.8:1	5.4:1	10.6:1	18.9:1	26.3:1

Figure 20 shows the spatial and spectral differences between the original and the FLEX compressed data for the cube 1732875312_1. We choose the bands at wavelengths of $2.0\,\mu\mathrm{m}$ and $5.0\,\mu\mathrm{m}$ where we can see the surface of Titan through its thick atmosphere. Three error values (e_1, e_2 and e_3) are applied with FLEX to assess the impact of lossy compression as compression ratio increases.At shorter wavelength ($2.0\,\mu\mathrm{m}$), where the signal-to-noise ratio (SNR) is high, the spatial comparison shows a small difference between the original and the decompressed data with maximum error e_1. We do notice that the spatial resolution of the image degrades when the compression ratio increases. The local micro-contrast decreases and some of the fine details on the surface are erased at high compression ratio (e_3) but the overall image is well reproduced.

At longer wavelength ($5.0\,\mu\mathrm{m}$), where the instrumental SNR is known to be lower,[3] the degradation of the image increases faster than at $2.0\,\mu\mathrm{m}$ as compression ratio increases. We also observe the emergence of horizontal striping effects due to the fact the decompressed data have the same intensity as the background noise (not compressed in our case), and the compression error leads to more detrimental effects compared to high SNR bands.

To illustrate this spectral dependency, we extracted the spectrum of a bright pixel at (9, 32) in the cube. All the windows and wings below $3\,\mu\mathrm{m}$ are correctly reproduced with an absolute difference smaller than 0.01 when compared with the original data. For longer wavelengths (i.e., larger than $3\,\mu\mathrm{m}$), this difference increases significantly due to the lower SNR associated with these bands.[3] To alleviate this spectral dependency of compression impacts, we changed the maximum error parameter so that it adapts to the signal level variations over different wavelengths. For a given targeted compression ratio, FLEX computes the band-dependent e so that a smaller error is used for low SNR bands.

In order to evaluate and quantify the effects of lossy compression for all of the selected VIMS cubes, we developed a metric to compare the reconstruction errors of lossy compression with the intrinsic deviation of the original data. The intrinsic deviation corresponds to the difference between the original data and the median value of the 3×3 surrounding pixels. This gives us an estimate of the expected variability of the data, i.e., shot noise, that can be used as a reference for tolerable compression error. We show in Figure 21 the scatter plots of the compression error versus the intrinsic noise for each of the selected cubes.

For a $64 \times 64 \times 256$ cube, we have more than 1 million points of comparison for any level of lossy compression. Visualizing this data-set is challenging and leads to saturation effect around the origin (0, 0). We do not use a density plot on purpose in order to showcase the change of the scatter patterns as compression ratio increases. We draw ellipses on top of the scatter plots that contain $68\,\%$, $95\,\%$ and $99.7\,\%$ of the data points. Each ellipse encircles the points whose compression error is below the corresponding percentile value. If the distribution of the compression error is Gaussian, this would represent $1\,\sigma$, $2\,\sigma$ and $3\,\sigma$ for the dispersion of our data-set. Therefore, an elongated ellipse along the x-axis corresponds to the case where reconstruction errors are small relative to the intrinsic noise of the cube; and an ellipse elongated along the y-axis correspond to the scenario where the reconstruction errors are larger than the intrinsic noise. We stopped increasing compression ratios when the ellipses start to elongate along the y-axis.As expected, for low compression levels, the reconstruction errors are smaller than the intrinsic deviations (ellipses elongated along the x-axis). The higher the compression ratio goes, the larger the y-axis of the ellipses become. The shape of the ellipses associated with lower percentiles change orientation faster than the ellipse of higher percentile. This means as compression ratio increases, the relative darker pixels are impacted more than the brighter signals. At high compression ratio (e.g., $\sim$20 :1),

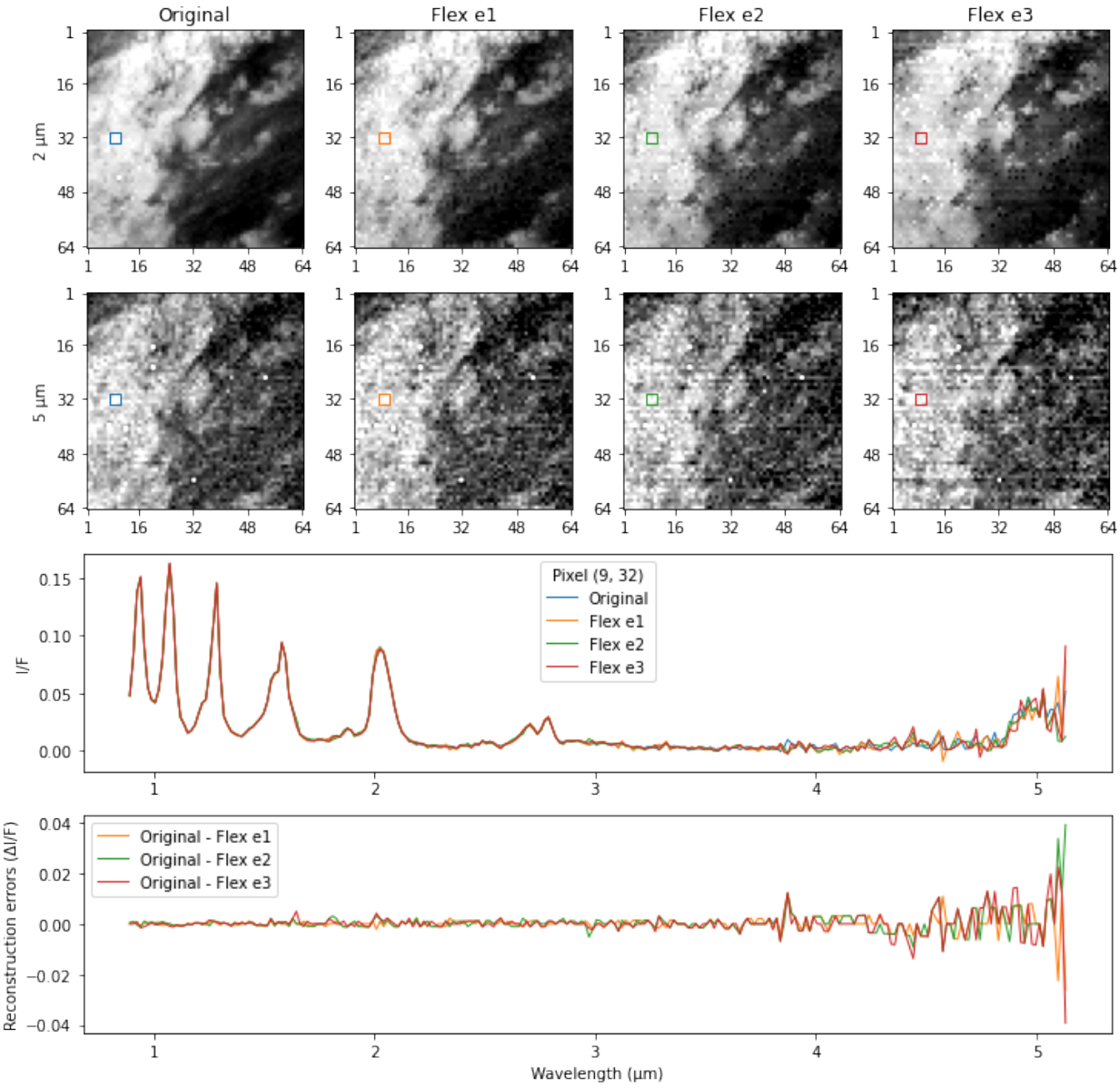


Fig 20 Comparison for the cube 1732875312_1 between the original data and the lossy compressed data with FLEX (CR from 15 :1 to 47 :1). The top two rows show the original and reconstructed images at 2.0 µm and 5.0 µm. In each image, we extracted the spectrum of the pixel (9, 32) highlighted by a colored square. The I/F spectrum and the difference between the original and the reconstructed spectrum are reported in the bottom two rows.

the ellipses are all elongated along the y-axis, implicating that the compression error is larger than the intrinsic noise for pixels of all brightness levels.

The extent to which reconstruction errors larger than the intrinsic noise (here $1\,\sigma$, $2\,\sigma$ or $3\,\sigma$ criteria) can be tolerated has to be defined by the science team with prior knowledge of the instrument response. For example, if we choose the compression ratio to ensure that all the $68\,\%$ of pixels with lower signal level can be reconstructed with error lower than their intrinsic noise, then the highest compression ratios are $\sim$7.8 :1, $\sim$10 :1, and $\sim$10 :1 for the high resolution, medium

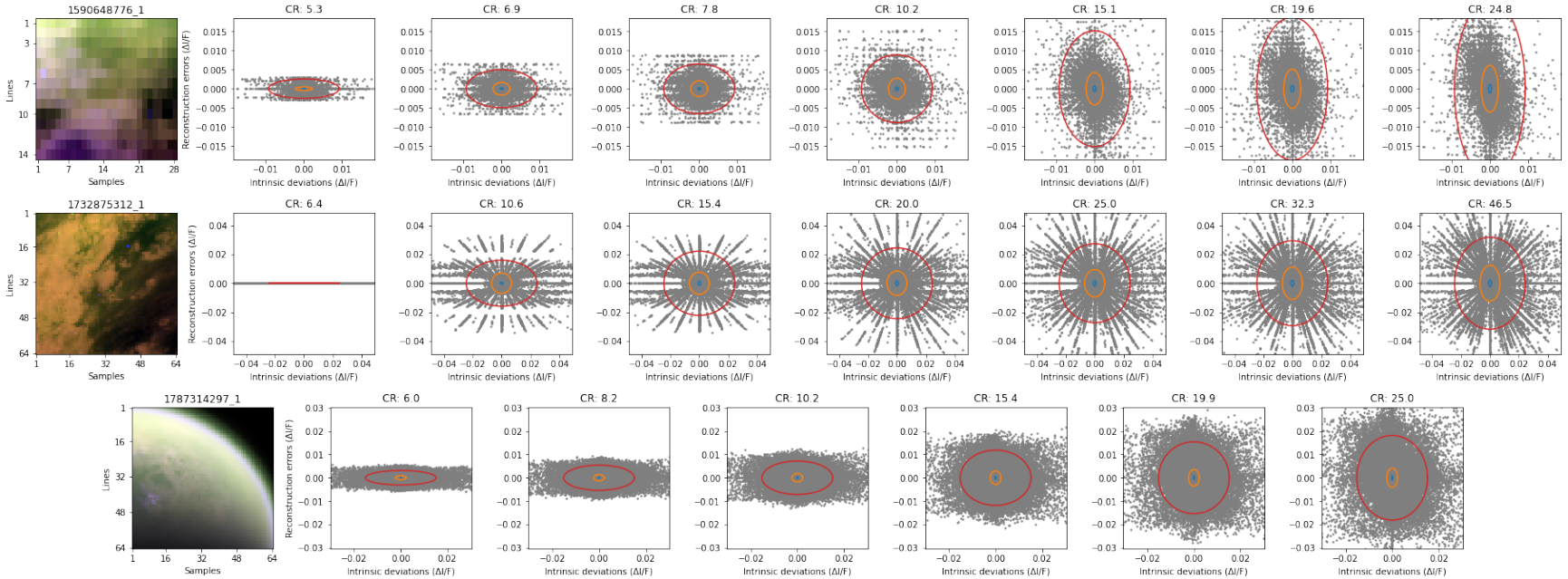


Fig 21 Comparison of compression error with intrinsic noise at increasing compression ratios for a high resolution cube (1590648776_1), a medium resolution cube (1732875312_1), and a low resolution cube (1787314297_1). The images on the left are RGB composites of Titan's surface at 2.0 µm, 1.6 µm and 2.8 µm. The gray scatter plots on the right represent compression error (y-axis) versus estimate of intrinsic noise (x-axis) for all pixels. The blue, orange and red ellipses contain 68 %, 95 % and 99.7 % of the dots respectively (1 σ, 2 σ and 3 σ criteria).

resolution, and low resolution cubes, respectively. On the other hand, higher compression ratios, e.g., $\sim$10 :1, $\sim$15 :1, and $\sim$20 :1, can be achieved when the tolerable error is chosen based on the 99.7 percentile of the intrinsic noise in the data. In the first case, it means that we put a high constraint on the fidelity of the reconstructed data to accurately replicate both the original signal and the noise of the instrument because the latter is excluded in the remaining 32 %. In the other case, when we relax the constraint, we allow the reconstructed data to introduce its own noise but with an amplitude lower than the noisiest part of the instrument. Therefore, it will be harder to get an accurate reconstruction of the data (with a elongated ellipse along the x-axis) with the $1\,\sigma$ criteria than with the $3\,\sigma$.

We extend this metric for all selected images (21 cubes). For medium resolution cubes such as 1732875312_1, we confirm that medium compression ratio (15.4 :1) led to reconstruction error lower than the intrinsic deviation of the data. However this is no longer the case for high compression ratios(e.g., 46.5 :1) as shown in the Figure 20. Moreover, we can see that the dispersion pattern is mainly driven by the digitalization of the data. For the low resolution cube (like 1787314297_1), where we can see the limb and/or the terminator, the intrinsic deviation become very broad and often larger than the reconstruction errors. This will allow a good reconstruction of the data even at high compression ratios (e.g., 19.9 :1).

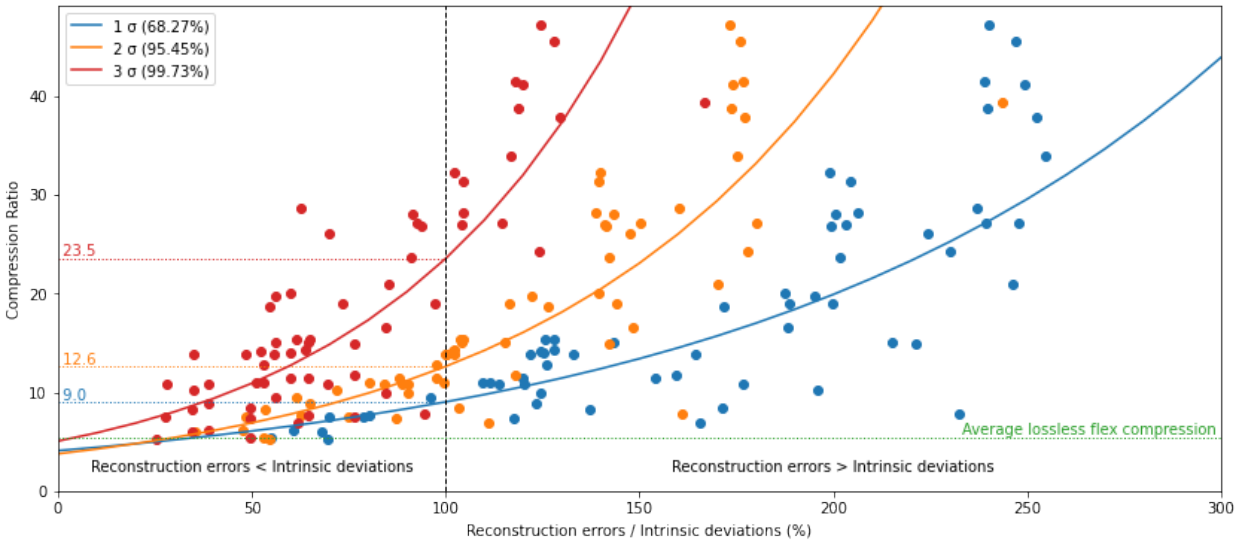


Fig 22 Relationship between the quality of the lossy reconstructed data and the compression ratio. The blue, orange and red dots corresponds to the $1\,\sigma$, $2\,\sigma$ and $3\,\sigma$ of the dispersion of the data from the Figure 21.

In general, for each compression ratio tested we can derive an estimate of the quality of the reconstructed data by dividing the y-axis of the ellipse by its x-axis. Based on the subset that we analyzed, we concluded that the data are correctly reproduced when the reconstruction errors are smaller than the intrinsic deviations. In Figure 22, this threshold is represented by the 100 % vertical dash line (i.e., a round circle in the Figure 21). Depending on the fidelity we want to achieve ($1\,\sigma$, $2\,\sigma$ or $3\,\sigma$), we can derive an empirical exponential law between the quality of the lossy reconstructed data and the optimal compression ratio at this threshold. This gives us an upper limit for the compression ratio after which the data will loose significant information.

Even with a restrictive constraint ($1\,\sigma$), we end up with an additional factor 2 of improvement in compression ($9.0:1$) comparing to the average lossless compression ratio with FLEX (e.g., $5.4:1$), without introducing too much error in the reconstructed data. In the less restrictive case ($3\,\sigma$), the compression ratio can be extended up to another factor 2.5 ($23.5:1$) in the best case scenarios with compression error below intrinsic noise level.

6 Scientific Distortion Metric

Our ultimate goal is to demonstrate state-of-the-art compression algorithms for future outer solar system space missions able to directly incorporate scientifically motivated distortion metrics in the compression scheme. This will enable us to develop a framework to generalize the classical Shannon rate-distortion theory to optimize scientific retrieval performance of the reproduced data.

As an example of the kind of distortion metric we have in mind, we focus on a particular example, spectral-imaging data, returning a spectrum at each pixel. Focusing on a single pixel, the scientific use of spectral imaging is often to retrieve physical parameters, for example the surface albedo. If we have a detailed physical model, for example a forward radiative transport model, predicting the observed spectrum as a function of the physical parameters, we can solve the inverse problem of finding the physical parameter which leads to a *predicted spectrum* that agrees with the *observed spectrum*. For an observed single-pixel spectrum $Y(\nu)$ and a predicted spectrum $F(\Theta)$ given by the physical parameters of interest Θ, an optimal estimation retrieval is given by the *Bayes maximum a posteriori estimator*,

$$\begin{aligned} \hat{\Theta}(Y) &= \min_{\Theta} \{-\log p[\Theta|Y]\} \\ &\sim (Y - F(\Theta))^{\dagger} N^{-1} (Y - F(\Theta)) + (\Theta - \Theta_0)^T S^{-1} (\Theta - \Theta_0) \end{aligned} \tag{2}$$

where $F(\Theta)$ is the forward radiative transport model spectrum, N^{-1} is the instrument *inverse-noise covariance matrix for the noise in each spectral bin*, and S^{-1} is the inverse physical parameter prior covariance matrix about a typical value Θ_0. For the present discussion, we are greatly simplifying the various choices of prior on the physical parameters that may be appropriate, and simply using the commonly used Gaussian prior in optimal estimation.

For the purposes of discussion and to motivate the emergence of a scientifically motivated distortion metric, we now approximate the Bayes parameter posterior with a Gaussian form

$$-\log p[\xi|Y] \sim (Y - F(\Theta) - J(\Theta_0)\xi)^{\dagger} N^{-1} (Y - F(\Theta_0) - J(\Theta_0)\xi) + \xi S^{-1} \xi \tag{3}$$

where the posterior is now in the physical parameter *variations about the prior mean* $\Theta = \Theta_0 + \xi$, and we have used a first order variation of the model-predicted spectrum $F(\Theta_0 + \xi) \approx F(\Theta_0) + J(\Theta_0)\xi$, with $J(\Theta_0)$ the model Jacobian, relating variations in the predicted spectrum with variations in the physical parameters (for K spectral bins and M physical parameters the Jacobian is an $K \times M$ matrix evaluated for the prior mean parameters Θ_0).

We pause to note that the Gaussian approximation we are making here will not be accurate unless both $\Theta_0 \approx \hat{\Theta}(Y)$, *and the instrument measurement noise is low enough that we are strongly data dominated*. When the latter condition is not true, the nonlinear nature of the forward physical model can render a Gaussian approximation to the Bayes parameter posterior inaccurate. Leaving this detail aside for the present discussion, we note that *computing the optimal estimate* $\hat{\Theta}(Y)$ *for detailed physical models is far too expensive to be computed on-board the spacecraft.* However, we can study the expense in memory incurred in *precomputing and storing a codebook*

$\mathcal{C} = \cup_k \{\Theta_k, F(\Theta_k), J(\Theta_k)\}$ of forward predicted spectra and their Jacobians in order to first classify the observed spectrum

$$\tilde{\Theta}(Y) \;=\; \min_{\Theta \in \mathcal{C}} \left\{ (Y - F(\Theta))^\dagger N^{-1} (Y - F(\Theta)) \right\} \tag{4}$$

For the remaining discussion, we assume some computationally inexpensive classification of the observed spectrum has been performed, and we use a Gaussian approximation about that codebook parameter, i.e.

$$\begin{aligned} -\log p[\Theta|Y] &\approx -\log p[\tilde{\Theta}(Y) + \xi | Y] \\ &\sim (Y - F(\tilde{\Theta}) - J(\tilde{\Theta})\xi)^\dagger N^{-1} (Y - F(\tilde{\Theta}) - J(\tilde{\Theta})\xi) + \xi S^{-1} \xi \end{aligned} \tag{5}$$

(where the Jacobians for the codebook have been precomputed $J(\tilde{\Theta})$, and we now drop the 'Y'-dependence and remember $\tilde{\Theta}$ is the optimal codebook physical parameter for the data Y).

We now return to the *transform coding* setting and assume a dictionary of basis functions, denoted W, and also assume that the dictionary is complete for the appropriate function space where our observed spectra Y live, so that with enough basis functions, we can achieve lossless representations of the data $Y = W\mathbf{a}_\infty$. For a chosen basis function dictionary, we therefore have the physical parameter Bayes posterior

$$-\log p[\xi|\mathbf{a}_\infty] \;\sim\; (W\mathbf{a}_\infty - F(\tilde{\Theta}) - J(\tilde{\Theta})\xi)^\dagger N^{-1} (W\mathbf{a}_\infty - F(\tilde{\Theta}) - J(\tilde{\Theta})\xi) + \xi S^{-1} \xi \tag{6}$$

Our goal is to quantify distortion in compression along scientifically meaningful directions, i.e. the Jacobian eigenmodes with large eigenvalues, and throw away the bits required to encode small informative content directions. If we consider another dictionary coefficient vector, such as the *coefficient thresholded version of* $\mathbf{a}_\infty$, we have a different posterior $-\log p[\xi|\mathbf{b}]$. A natural way to quantify the "significance" of replacing the lossless representation of the data Y with the uncompressed dictionary coefficient vector $\mathbf{a}_\infty$ with a compressed version $\mathbf{b}$ is with the Kullback-Leibler distortion between the two Bayes posteriors. A direct computation in the Gaussian approximation to our Bayes posteriors gives

$$\begin{aligned} J(\mathbf{b}, \mathbf{a}_\infty) &= \int d[\xi] \;\; (p[\xi|\mathbf{b}] - p[\xi|\mathbf{a}_\infty]) \log \left(\frac{p[\xi|\mathbf{b}]}{p[\xi|\mathbf{a}_\infty]} \right) \\ &= \left[\hat{\xi}(\mathbf{b}) - \hat{\xi}(\mathbf{a}_\infty) \right]^\dagger \left[J^T N^{-1} J + S^{-1} \right] \left[\hat{\xi}(\mathbf{b}) - \hat{\xi}(\mathbf{a}_\infty) \right] \end{aligned}$$

where we defined

$$\hat{\xi}(\mathbf{b}) \;=\; \left[J^T N^{-1} J + S^{-1} \right]^{-1} J^T N^{-1} \left(W\mathbf{b} - F(\tilde{\Theta}) \right) \tag{7}$$

Assuming that $Y = W\mathbf{a}_\infty$, and defining the positive definite matrix

$$\mathcal{M} \;=\; \left\{ N^{-1/2} J S^{1/2} \left[I + S^{1/2} J^T N^{-1} J S^{1/2} \right]^{-1} S^{1/2} J^T N^{-1/2} \right\} \tag{8}$$

this gives

$$J(\mathbf{b}, \mathbf{a}_\infty) \;=\; (Y - W\mathbf{b})^\dagger \, N^{-1/2} \mathcal{M} N^{-1/2} \, (Y - W\mathbf{b}) \tag{9}$$

The positive-definite matrix in $\mathcal{M} \equiv \{\cdot\}$ represents an effective projection of deviations in the compressed to uncompressed spectrum onto the subspace spanned by the *scientifically important eigenmodes with appreciable measurement signal-to-noise*. Directions in the basis function dictionary W along non-informative or poorly measured directions can be thresholded to zero. Our scientifically motivated objective for compression can then be generalized to

$$\begin{aligned} \hat{b} &= \min_{\mathbf{b}} \{J(\mathbf{b}, Y) + \lambda \|\mathbf{b}\|^p\} \\ &= \min_{\mathbf{b}} \left\{ (Y - W\mathbf{b})^\dagger N^{-1/2} \mathcal{M} N^{-1/2} (Y - W\mathbf{b}) + \lambda \|\mathbf{b}\|^p \right\} \end{aligned} \tag{10}$$

Ideally we could bring this kind of scientific insight into the compression algorithm, tailoring *the measure of distortion it to the physics of the situation*.

The expression in equation 10 both suggests physically motivated dictionaries for compression, as well as generalizations of the above to allow for lossless compression *when we are surprised by the observations in the context of the assumed physical theory*. For the former, equation 10 suggests that the most efficient set of basis functions would *be the eigenvectors of* $\mathcal{M}$, and intuitively is closely related to our SVD based-compression in the context of a physically motivated theoretical framework within which we interpret the observations (SVD as presented above can be thought of as an empirical determination of the optimal dictionary without the background context of a physical model). However, other families of dictionaries such as the wave atoms considered in this paper provide the interesting compromise of efficient compression for observations that are *surprising* with respect to the physical model, and more efficient at compression (in an appropriately chosen background function space reflecting signal smoothness properties, etc.) with respect to the standard least squares error distortion.

It is interesting to consider therefore a one-parameter family of compression objective functions for some $\tau \in [0, 1/2]$

$$\hat{b} = \min_{\mathbf{b}} \left\{ (Y - W\mathbf{b})^\dagger N^{-1/2} \left[(1-\tau)\mathcal{M} + \tau(\mathcal{I} - \mathcal{M}) \right] N^{-1/2} (Y - W\mathbf{b}) + \lambda \|\mathbf{b}\|^p \right\} \tag{11}$$

where $\mathcal{I}$ = Identity and $\mathcal{I} - \mathcal{M}$ represents a "projection" of the data into directions that the physical model would "ignore" as physically uninformative. It is exactly those directions which potentially contain "surprises" and scientific discoveries as well noise, so care should be taken in allocating bits to this function subspace. If an automatic trigger of "surprise" was available upon receiving new observations Y, we could allow $\tau \to 1/2$, returning the compression scheme back to the standard minimum least squares transform coding algorithm. It is interesting to consider the development of algorithms to *automatically classify an observation as surprising from the point of view of the physical model and react by adjusting* τ. Future work will develop such dynamic compression schemes, using either Bayes optimal estimates of τ or simple discrete thresholding on $\tau \in \{0\} \cup \{1/2\}$ depending on the physical model ensemble distribution of a test statistic $\Psi = (Y - W\hat{b})N^{-1/2}(\mathcal{I} - \mathcal{M})N^{-1/2}(Y - W\hat{b})$. For received observations Y and fixed value of λ, the $\tau = 0$ transform coding solution can be found $\hat{b}_0$, followed by the evaluation of $\Psi[\hat{b}_0]$, which is a measure of the "energy" in Y not accounted for in the physical model subspace. Our surprise at this quantify can be related to the distribution under the physical model (and noise properties of the instrument), $P_\lambda[\Psi \geq \Psi(\hat{b}_0)]$. A simple binary threshold decision can be made, for example for some γ such that $P_\lambda[\Psi \geq \Psi(\hat{b}_0)] \geq \gamma \Rightarrow \tau = 0$, and $\tau = 1/2$ otherwise. If there are unexpected features missing in the compressed data with $\tau = 0$, then $\Psi(\hat{b}_0)$ will be

larger than expected under the physical model prior and instrument noise ensemble, meaning the probability of such a large deviation is smaller than the chosen threshold γ, triggering a switch back to the inverse-noise weighted square error compression problem. Future work will explore optimal choices and consequences of this strategy, and compare with soft-thresholding, directly relating τ to the large-deviation theory rate functions, and possible Bayes optimal approaches to fixing τ.

Exploring the ideas above for science motivated distortion metrics was beyond the scope of this paper, as it involves the computationally demanding merger of detailed forward physical models and optimal estimation with compression. Future work will endeavor to do just that, and again rely on feedback from the science community on tailoring the "gains" in the dynamic allocation of bits capturing scientific surprise.

7 Conclusions and Future Work

In this retrospective study we investigated data compression algorithms to boost science data return from high data volume instruments on planetary missions, particularly outer solar system missions. We developed a methodology for development of advanced compression with iteration and evolutionary improvement via science team evaluation in real time or periodically. We show that an iterative process involving real-time science team evaluation and feedback to update the on-board compression algorithm is both essential and feasible. Beyond the obvious benefits of increased science return and faster playback, better data compression enables design trades between antenna size and number of science instruments on the payload.

In the next phase of our study, we will augment compression algorithms with machine learning to discover improvements in the encoder-decoder that can in principle achieve the Shannon rate-distortion theoretic ideal. Using the established methodology in this study, we will demonstrate and quantify the impact on science with compression algorithms optimized with respect to novel, scientifically-motivated measures of distortion. The long-term goal is to infuse into future missions the techniques we develop and to provide a path for science teams to evaluate and to feel comfortable using more aggressive compression.

Appendices

A Wave Atoms

Wave atoms are constructed from tensor products of adequately chosen wavelet packets. Simple, real-valued wave atoms in the 1D case can be obtained by combining dyadic dilates and translates of the mother wave atom $\hat{\psi}_m^0$ on the frequency axis. An elegant methodology for constructing $\hat{\psi}_m^0$ was given by Lars Villemoes[33] to resolve the frequency localization problem associated with standard wavelet packet analysis. Equation 12 gives the Villemoes construction of mother wave atom:

$$\hat{\psi}_m^0(\omega) = e^{-i\omega/2}\left[e^{i\alpha_m} g\left(\epsilon_m(\omega - \pi(m+0.5))\right) + e^{-i\alpha_m} g\left(\epsilon_{m+1}(\omega + \pi(m+0.5))\right)\right], \tag{12}$$

where $\alpha_m = \pi/2(m + 0.5)$ and $\epsilon_m = (-1)^m$. The function g is an real-valued, C^∞ bump function with compact support on an interval of length 2π such that

$$\sum_m |\hat{\psi}^0_m(\omega)|^2 = 1.$$

The wave atom basis function at scale j, centered in frequency around $\pm\omega_{j,m} = \pm\pi 2^j m$ and in space around $x_{j,n} = 2^{-j}n$ can be obtained as

$$\psi^j_{m,n}(x) = 2^{j/2}\mathcal{F}^{-1}\left(\hat{\psi}^0_m\right)\left(2^j x - n\right), \tag{13}$$

where $j \geq 0$, $C_1 2^j \leq m \leq C_2 2^j$, and $\mathcal{F}^{-1}$ is the inverse Fourier transform.

The orthonormality of the wave atom basis functions can be preserved via an asymmetry support condition on g, i.e.,

$$\forall\omega \in \left[\frac{-\pi}{3}, \frac{\pi}{3}\right] : g\left(-2\omega - \frac{\pi}{2}\right) = g\left(\omega + \frac{\pi}{2}\right). \tag{14}$$

Typical candidates for g are based on local cosine bases. We used the implementation provided in the wave atom Matlab toolbox.[29] While various variants of the transform can be obtained by splitting up the bumps into individual atoms, we used the real-valued representation with 2^d (d is the number of dimensions of the data) bumps per wave atom since it provides us with a critically sampled representation.

The wavelet packet tree, defining the partitioning of the frequency axis in 1D, is shown in Figure 23.

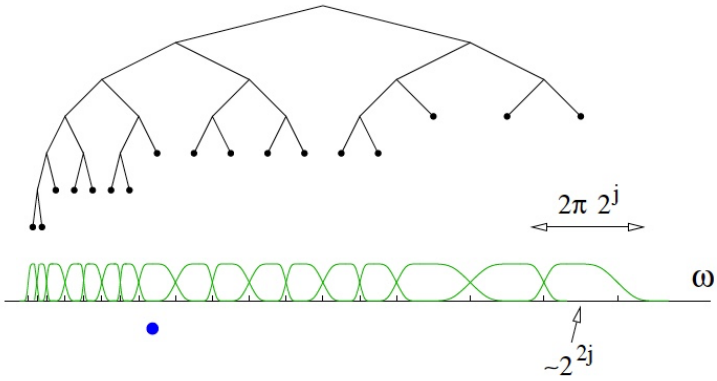


Fig 23 The wavelet-packet tree corresponding to wave atoms. The bottom graph demonstrates Villemoes[33] wave atoms on the positive frequency axis.

Acknowledgments

We would like to thank Aaron Kiely and Matthew Klimesh for providing us the software for FLEX compression and Rice compression. The research described in this publication was carried out in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

References

1 C. C. Porco, R. A. West, S. Squyres, *et al.*, "Cassini Imaging Science: Instrument Characteristics and Anticipated Scientific Investigations At Saturn," *Space Sci. Rev.* **115**(1-4), 363–497 (2004).

2 D. A. Gurnett, W. S. Kurth, D. Kirchner, *et al.*, *The Cassini Radio and Plasma Wave Investigation*, Springer (2004).

3 R. H. Brown, K. H. Baines, G. Bellucci, *et al.*, "The Cassini Visual and Infrared Mapping Spectrometer (VIMS) Investigation," in *The Cassini-Huygens Mission*, , Ed., 111–168, Kluwer Academic Publishers, Dordrecht (2004).

4 L. W. Esposito, C. A. Barth, J. E. Colwell, *et al.*, *The Cassini Ultraviolet Imaging Spectrograph Investigation*, Springer (2004).

5 L. Woolliscroft, W. Farrell, H. ALLEYNE, *et al.*, "Cassini Radio and Plasma Wave Investigation: Data Compression and Scientific Applications," *British Interplanetary Society, Journal* **46**(3), 115–120 (1993).

6 I. Willis, L. Woolliscroft, T. Averkamp, *et al.*, "The Implementation of Data Compression in the Cassini RPWS Dedicated Compression Processor," in *Proceedings DCC'95 Data Compression Conference*, 496–496, IEEE Computer Society (1995).

7 R. Rice, P. Yeh, and W. Miller, "Algorithms for a Very-High-Speed, Noiseless Encoding Module," tech. rep., Jet Propulsion Laboratory (1991).

8 P. Smith, J. Bell, N. Bridges, *et al.*, "Results from the Mars Pathfinder Camera," *Science* **278**(5344), 1758–1765 (1997).

9 A. Kiely and M. Klimesh, "Preliminary Image Compression Results From the Mars Exploration Rovers," *IPN Progress Report* **42**, 156 (2004).

10 Recommendation for Space Data System Standards, *Image Data Compression*, CCSDS 122.0-B-1, Blue Book, Washington, D.C. (2005).

11 A. Kiely and M. Klimesh, "The ICER Progressive Wavelet Image Compressor," *Interplanetary Network Progress Report* **42-155**, 1–46 (2003).

12 "JPEG2000 Image Coding System: Extensions." ISO/IEC 15444-1.

13 L. Demanet and L. Ying, "Wave Atoms and Sparsity of Oscillatory Patterns," *Appl. Comput. Harmon. Anal* **23** (2006).

14 A. M. Rufai, G. Anbarjafari, and H. Demirel, "Lossy Image Compression Using Singular Value Decomposition and Wavelet Difference Reduction," *Digital signal processing* **24**, 117–123 (2014).

15 I. Daubechies, *Ten Lectures On Wavelets*, vol. 61, Siam (1992).

16 R. R. Coifman, Y. Meyer, S. Quake, *et al.*, "Signal Processing and Compression with Wavelet Packets," in *Wavelets and Their Applications*, , Ed., 363–379, Springer Netherlands (1994).

17 C. Herley, J. Kovacevic, K. Ramchandran, *et al.*, "Tilings of The Time-frequency Plane: Construction of Arbitrary Orthogonal Bases and Fast Tiling Algorithms," *IEEE Transactions on Signal Processing* **41**(12), 3341–3359 (1993).

18 D. L. Donoho, M. Vetterli, R. A. DeVore, *et al.*, "Data Compression and Harmonic Analysis," *IEEE Transactions on Information Theory* **44**(6), 2435–2476 (1998).

19 R. Kumar, A. Kumar, and G. Singh, "Hybrid Method Based on Singular Value Decomposition and Embedded Zero Tree Wavelet Technique for ECG Signal Compression," *Computer Methods and Programs in Biomedicine* **129**, 135–148 (2016).

20 J. C. S. de Souza, T. M. L. Assis, and B. C. Pal, "Data Compression in Smart Distribution Systems via Singular Value Decomposition," *IEEE Transactions on Smart Grid* **8**(1), 275–284 (2015).

21 S. K. Mukhopadhyay, M. O. Ahmad, and M. Swamy, "An ECG Compression Algorithm with Guaranteed Reconstruction Quality Based on Optimum Truncation of Singular Values and ASCII Character Encoding," *Biomedical Signal Processing and Control* **44**, 288–306 (2018).

22 J. Guo, R. Xie, and G. Jin, "An Efficient Method for NMR Data Compression Based on Fast Singular Value Decomposition," *IEEE Geoscience and Remote Sensing Letters* **16**, 301–305 (2019).

23 A. Gersho, "Adaptive Filtering with Binary Reinforcement," *IEEE Transactions on Information Theory* **30**(2), 191–199 (1984).

24 B. Widrow, J. R. Glover, J. M. McCool, *et al.*, "Adaptive Noise Cancelling: Principles and Applications," *Proceedings of the IEEE* **63**(12), 1692–1716 (1975).

25 M. Klimesh, "Low-Complexity Lossless Compression of Hyperspectral Imagery via Adaptive Filtering," *Interplanetary Network Progress Report* **163**, 1–10 (2005).

26 Recommendation for Space Data System Standards, *Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression*, CCSDS 123.0-B-2, Blue Book, Washington, D.C. (2019).

27 R. N. Clark, R. H. Brown, D. M. Lytle, *et al.*, "The VIMS Wavelength and Radiometric Calibration 19, Final Report," tech. rep., The Planetary Atmospheres Node (2018).

28 J. Pickett, W. Kurth, D. Gurnett, *et al.*, "Electrostatic Solitary Waves Observed at Saturn by Cassini inside 10 Rs and Near Enceladus," *Journal of Geophysical Research: Space Physics* **120**(8), 6569–6580 (2015).

29 L. Demanet and L. Ying, "Wave Atom Repository." http://waveatom.org/.

30 T. K. et al., "The Mesosphere and Thermosphere of Titan Revealed by Cassini/UVIS Stellar Occultations," *Icarus* **216**, 507–534 (2011).

31 C. R. Wu, F. Chen, and D. Judge, "Measurements of Temperature-Dependent Absorption Cross Sections of C2H2 in the VUV-UV Region," *Journal of Geophysical Research: Planets* **106**(E4), 7629–7636 (2001).

32 S. Le Mouélic, T. Cornet, S. Rodriguez, *et al.*, "The Cassini VIMS Archive of Titan: From Browse Products to Global Infrared Color Maps," *Icarus* **319**, 121–132 (2019).

33 L. Villemoes, "Wavelet Packets with Uniform Time-Frequency Localization," *Comptes Rendus Mathematique - C R MATH* **335**, 793–796 (2002).