AIAA SciTech Forum
23-27 January 2023, National Harbor, MD & Online
AIAA SCITECH 2023 Forum

10.2514/6.2023-1684

Evaluating the Impact of Lossy Compression on a Direct Numerical Simulation of a Mach 2.5 Turbulent Boundary Layer

Christian J. Lagares-Nieves[1]* and Guillermo Araya[1]†
[1]*HPC and Visualization Lab, Dept. of Mechanical Eng., University of Puerto Rico at Mayaguez, PR 00681, USA.*

Lossy compression techniques are ubiquitous in many fields including imagery and video; however, the incursion of such lossy compression techniques in the computational fluid dynamics community has not advanced to the same extent in decades. In this work, the lossy compression of high-fidelity direct numerical simulation (DNS) is evaluated to assess the impact on various parameters of engineering interest. A Mach 2.5, spatially developing turbulent boundary layer (SDTBL) at a moderately high Reynolds number has been selected as the subject of the study [1]. The ZFP compression scheme was chosen as the core driving algorithm for this study as it was carefully crafted for scientific, floating point data. The resilience of spectral quantities as well as two-point correlations is highlighted. Notwithstanding, we also noted that point-wise values calculated in the physical domain were prone to quantization errors at high compression ratios. Further, we have also presented the impact on higher order statistics. In summary, we have demonstrated that high fidelity results are within reach while achieving 1.45x to 9.82x reductions in required storage over single precision, IEEE 754-compliant data values.

I. Nomenclature

U = Time-averaged streamwise velocity
V = Time-averaged wall-normal velocity
W = Time-averaged spanwise velocity
T = Time-averaged Temperature
P = Time-averaged Pressure
M_∞ = Freestream Mach Number
$Re_{\delta 2}$ = Momentum thickness Reynolds Number
u_τ = Friction velocity
δ = Velocity boundary layer thickness
δ_1 = Displacement thickness
δ_2 = Momentum thickness
ρ = Fluid density
τ_w = Wall shear stress
$(\cdot)'$ = Fluctuation with respect to time-averaged mean
$(\cdot)^+$ = Inner-scaled quantity
$(\cdot)_\infty$ = Quantity measured or calculated at freestream conditions
$(\cdot)_w$ = Quantity measured or calculated at the wall surface
GB = Gigabyte

II. Introduction

The renaissance of high-speed aerodynamics, space exploration and military technology advances imposes a strong weight on the computational fluid dynamics (CFD) community to achieve high fidelity results at higher Reynolds numbers and evermore complex geometries. Although computational performance has grown exponentially over the

---

[*] PhD Research Assistant & Doctoral Candidate, Department of Mechanical Engineering, 259 Alfonso Valdez Blvd., Mayaguez, PR, 00680. AIAA Student Member, christian.lagares@upr.edu

[†] Associate Professor, Department of Mechanical Engineering, 259 Alfonso Valdez Blvd., Mayaguez, PR, 00680. AIAA Associate Fellow, araya@mailaps.org

years, data storage has lagged far behind in terms of raw bandwidth and raw capacity. Storing large CFD simulations for posterity and future analysis/comparisons presents challenges that have been seen in other fields in the past. For instance, static imagery is often stored in compressed formats such as JPEG [2] and PNG [3]. The JPEG file format is an example of a lossy file format where some quality is loss and data is not exactly preserved [2]. The PNG file format performs lossless compression where the data stored is recovered without loss (assuming no hardware errors) [3]. However, the JPEG file formats disregards high-frequencies as these may be harder to perceive by the naked eye. Coming back to the scientific endeavour at hand, one cannot simply disregard high-frequency terms in CFD datasets nor the fact that the flow is spatially correlated. This presents enormous challenges for data archival since often times only lossless compression schemes are considered, and turbulence datasets are not usually prone to high compression ratios due to the complex nature of the data.

As early as in 1998, Wilson argued that turbulence statistics were not well preserved in the presence of lossy compression schemes available at the time [4]. Fast forward 20 years after that statement, much has changed in landscape. Lossy compression has slowly gained attention from researchers due to the ever-growing computational and storage demands. Otero *et al.* proposed a domain-specific method based on the discrete Legendre transform [5]. The results presented by Otero and others were extremely favorable and suggested lossy data compression techniques were maturing and were advancing at a pace that might make them suitable enough for production usage within the broader CFD community.

Notwithstanding, lossy compression of scientific datasets is under active research. Somewhat simpler techniques such as frequency downsampling, decimal rounding, bit grooming, bit shaving and many others lack an adaptive nature and were not developed with the floating point nature of scientific datasets in mind. Lindstrom proposed an alternative approach for scientific compression library that eventually became ZFP [6]. ZFP operates on 4^d blocks of hyper-spatial data where d is the data dimensionality. As such, it can account for spatial dependencies and adapt to local values. ZFP maps floating point data into a custom format that is often more accurate than the IEEE floating point standard. It allows for a variable bit-depth and fixed-accuracy (or vice versa), and its design preserves derived physical quantities including spectral properties with higher accuracy than other methods. ZFP is also extremely efficient at the hardware level making it an attractive alternative for extremely large scientific datasets. Aside from every benefit ZFP offers, it is capable of operating over an arbitrary number of dimensions and floating point data types which makes it extremely versatile regardless of the scientific data at hand.

In this work, we explore the impact of lossy compression on a dataset for a moderately high Reynolds number turbulent boundary layer at Mach 2.5. We will present the impact on the energy spectra and two-point correlations as an initial vantage point followed by a convergence analysis for point-wise, first-order results. Lastly, we will highlight the impact (or lack thereof) of the lossy compression scheme for higher-order statistics. In all, we will explore compression schemes of variable bit-depths and fixed accuracy ranging from 10^{-1} to within machine precision (i.e., lossless).

III. Numerical Details

A. Simulation Details

Unsteady numerical predictions of spatially-developing turbulent boundary layers (SDTBL) by using DNS is non-trivial since high mesh resolution is required in order to resolve the smallest turbulence scales (Kolmogorov and Batchelor scales for momentum and thermal transport, respectively). Additionally, the computational domain should be large enough to appropriately resolve the influence of the large scale motions (LSM), as described in Hutchins & Marusic [7], which are located in the outer region of the boundary layer. The last requirement, but by no means the least important, dictates that realistic time-dependent inflow turbulent conditions (instantaneous velocity, temperature and pressure) must be prescribed, Araya *et al.*, [8] [9], and more recently by [10]. As articulated earlier, one of the key aspects on the simulations of unsteady spatially-developing turbulent boundary layers is the prescription of accurate and physically sound turbulent inflow information. In this study, we are using the inflow generation method devised by Araya *et al.* [8], which is an improved version of the original rescaling-recycling method by Lund *et al.* [11], and it has recently been extended to compressible flows in [9] [10]. An infographic of the DMA methodology is displayed in fig. 1. The fundamental idea of the rescaling-recycling method is to extract the flow solution (mean and fluctuating components of the velocity, thermal and pressure fields) from a downstream plane (called “Recycle") and after performing a transformation by means of scaling functions, the transformed profiles are re-injected at the inlet plane, as seen in figure 2, where iso-surfaces of the instantaneous streamwise velocity (normalized by the freestream velocity) can be seen. The major objective of implementing scaling laws to the flow solution is to convert the streamwise in-homogeneity of the flow into quasi-homogeneous conditions. This inflow generation technique has demonstrated to produce inlet turbulent conditions with the appropriate power spectra of velocity fluctuations [12]. In fig. 3, contours of instantaneous density (top) and density gradient magnitude or kind of Schlieren image (bottom). The leftmost and rightmost planes represent the inflow and outflow conditions. As we can see, flow develops naturally with the presence of bulges and valleys, which is typical in the turbulence structure of boundary layers. The classical Reynolds decomposition is implemented for instantaneous flow parameters, i.e. a time-averaged plus a fluctuating component:

$$u_i(\mathbf{x}, t) = U_i(x, y) + u_i'(\mathbf{x}, t) \tag{1}$$

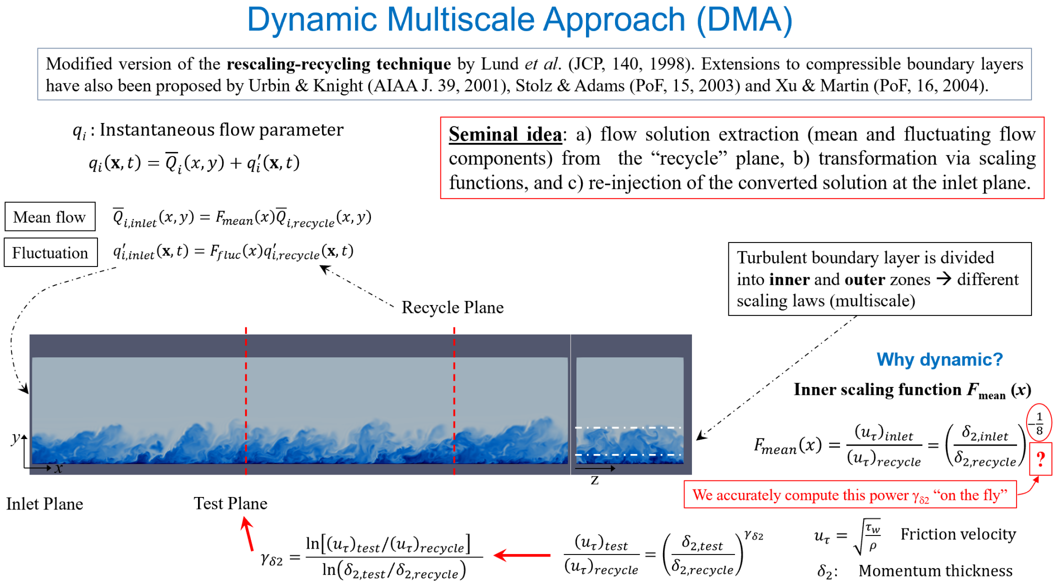
$$t(\mathbf{x}, t) = T(x, y) + t'(\mathbf{x}, t) \tag{2}$$

$$p(\mathbf{x}, t) = P(x, y) + p'(\mathbf{x}, t) \tag{3}$$

Furthermore, the turbulent boundary layer is divided into inner and outer zones, where different scaling laws are applied [8]. The projection of flow parameters from the recycle plane to the inlet is performed along constant values of y^+ (inner region) and y/δ (outer region). During the re-scaling process of the flow parameters in the inflow generation methodology [8], the ratio of the inlet friction velocity to the recycle friction velocity (i.e., $\lambda = u_{\tau,inl}/u_{\tau,rec}$) is required. The friction velocity is defined as $u_\tau = \sqrt{\tau_w/\rho}$, where τ_w is the wall shear stress and ρ is the fluid density. The inlet boundary layer thickness must be prescribed according to the desired inlet Reynolds number to be simulated, thus, prescribing also the inlet friction velocity would be redundant. To address this issue, Lund *et al.* [11], Urbin & Knight [13] and Stolz & Adams [14] used the well-known 1/8-power law that connects the friction velocity to the momentum thickness in zero-pressure gradient flows; thus, $u_{\tau,inl}/u_{\tau,rec} = (\delta_{2,inl}/\delta_{2,rec})^{-1/8}$. The empirical power (-1/8) is strongly affected by the Reynolds number dependency; therefore, we explicitly compute this power, $\gamma_{\delta 2}$, "on the fly" by relating the mean flow solution from a new plane (so-called the "Test" plane, as seen in figure 1) to the solution from the recycle plane as follows:

$$\gamma_{\delta 2} = \frac{ln(u_{\tau,test}/u_{\tau,rec})}{ln(\delta_{2,test}/\delta_{2,rec})}. \tag{4}$$

In order to carry out the proposed DNS, a highly accurate, very efficient, and highly scalable CFD solver is required. The flow solver PHASTA is an open-source, parallel, hierarchic (2^{nd} to 5^{th} order accurate), adaptive, stabilized (finite-element) transient analysis tool for the solution of compressible [15] or incompressible flows (Jansen [16]). PHASTA has been extensively validated in a suite of DNS under different external conditions [17], [18], [19], [10].



Dynamic Multiscale Approach (DMA)

Modified version of the **rescaling-recycling technique** by Lund *et al.* (JCP, 140, 1998). Extensions to compressible boundary layers have also been proposed by Urbin & Knight (AIAA J. 39, 2001), Stolz & Adams (PoF, 15, 2003) and Xu & Martin (PoF, 16, 2004).

q_i : Instantaneous flow parameter

$$q_i(\mathbf{x}, t) = \overline{Q}_i(x, y) + q_i'(\mathbf{x}, t)$$

Seminal idea: a) flow solution extraction (mean and fluctuating flow components) from the "recycle" plane, b) transformation via scaling functions, and c) re-injection of the converted solution at the inlet plane.

Mean flow $\overline{Q}_{i,inlet}(x, y) = F_{mean}(x)\overline{Q}_{i,recycle}(x, y)$

Fluctuation $q'_{i,inlet}(\mathbf{x}, t) = F_{fluc}(x)q'_{i,recycle}(\mathbf{x}, t)$

Turbulent boundary layer is divided into **inner** and **outer** zones → different scaling laws (multiscale)

Why dynamic?

Inner scaling function F_{mean} (x)

$$F_{mean}(x) = \frac{(u_\tau)_{inlet}}{(u_\tau)_{recycle}} = \left(\frac{\delta_{2,inlet}}{\delta_{2,recycle}}\right)^{-\frac{1}{8}} \boxed{?}$$

We accurately compute this power $\gamma_{\delta 2}$ "on the fly"

$$\gamma_{\delta 2} = \frac{\ln[(u_\tau)_{test}/(u_\tau)_{recycle}]}{\ln(\delta_{2,test}/\delta_{2,recycle})} \longleftarrow \frac{(u_\tau)_{test}}{(u_\tau)_{recycle}} = \left(\frac{\delta_{2,test}}{\delta_{2,recycle}}\right)^{\gamma_{\delta 2}}$$

$u_\tau = \sqrt{\frac{\tau_w}{\rho}}$ Friction velocity

δ_2: Momentum thickness

Fig. 1 The DMA rescaling-recycling method.

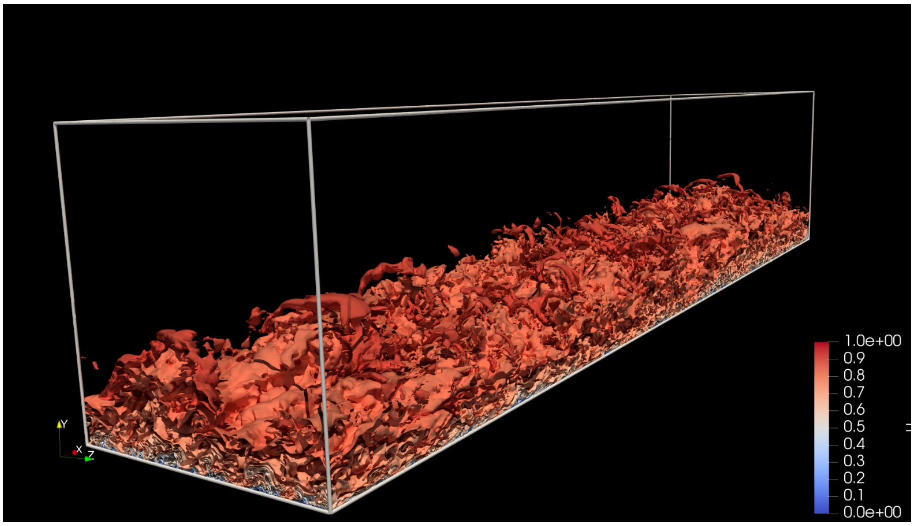


Fig. 2 Iso-surfaces of instantaneous streamwise velocity normalized by the freestream velocity.

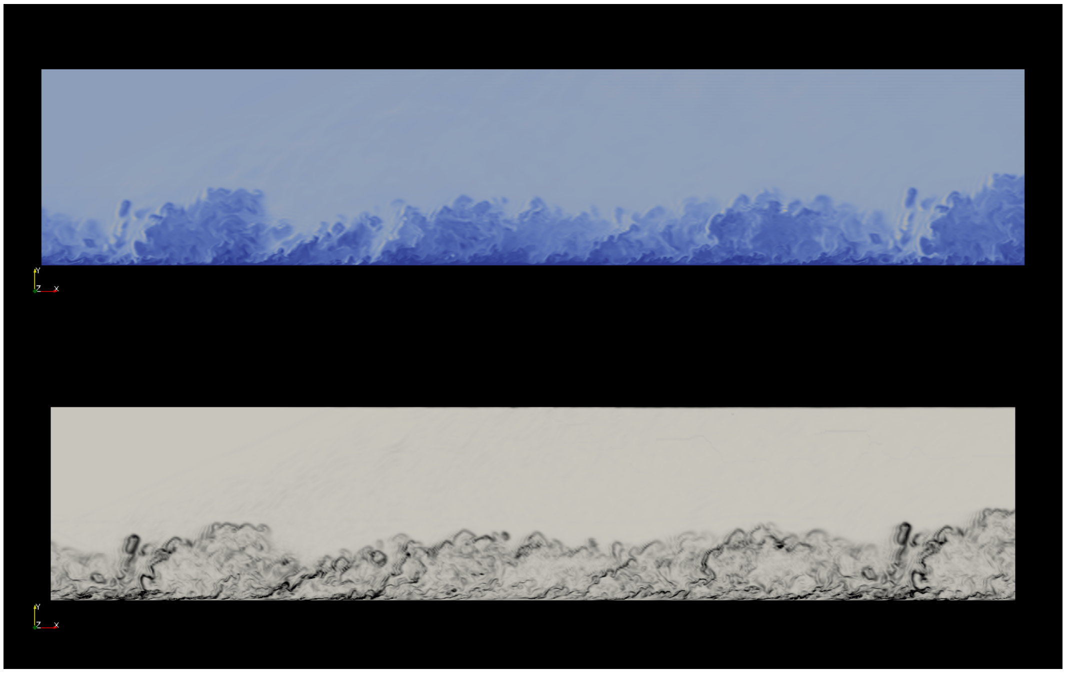


Fig. 3 Instantaneous density (top) and density gradient magnitude (bottom) in DNS of supersonic zero-pressure gradient boundary layers (M_∞ = 2.5) at a momentum thickness Reynolds number ($Re_{\delta 2}$) of 3,000 or $\delta^+ \approx 900$ (von Karman number).

Boundary Conditions: At the wall, the classical no-slip condition is imposed for all velocity components. Isothermal wall (quasi-adiabatic condition) is assumed for the thermal field. The ratio T_w/T_∞ is 2.25, where T_w is the wall temperature and T_∞ is the freestream temperature. The T_r/T_∞ ratio is 2.12 for M_∞ equals to 2.5. T_r is the recovery or adiabatic temperature. The lateral boundary conditions are handled via periodicity.

Validation of present DNS can be found in [1, 20]. Readers are referred to [1, 20] for numerical details as well as for more physics discussion.

B. Compression Scheme

As presented in [6], the compression scheme is segmented in five steps. We will briefly discuss these steps as the core mechanics of the compression scheme are not the focus of this work.

The first two steps align the floating point values within a 4^d-block to ensure a common exponent is shared. In essence, this yields a localized, adaptive fixed-point representation. This exponent is limited by the largest (absolute) value owned by the block. The next step performs a block transform that decorrelates the values within the block. Spatially correlated values (as is common with physics simulations) makes it much more difficult to efficiently compress data. This is achieved with per-dimension 1D transformations. ZFP includes a custom, efficient transformation scheme akin to the discrete cosine transform used by JPEG. The coefficients of the transform are finally aligned according to the expected magnitude and encoded on a per-bit-plane basis. Once again, we are vague in details because the actual compression scheme is far beyond the scope of this work. The interested reader is referred to [6] for further details on each step and the rational behind each one.

C. Post-Processing

The results contained in this work were obtained using a recent version of our in-house post-processing library, Aquila [21]. Aquila V3 targets TensorFlow [22] as a performance portability backend enabling transparent execution across CPUs and GPUs from multiple vendors and on multiple operating systems. We leveraged the ZFP [6] library interface via HDF5 [23].

IV. Results and Discussion

In this section, we will focus on the impact that lossy compression has over quantities of scientific and engineering interest. We will first tackle the gains in terms of storage footprint followed by the behavior of the compression scheme in spectral and two-point correlations since many prior methods were based on, derived from or inspired on frequency downsampling (as are other popular schemes in other fields such as JPEG [2]).

A. Dataset Size Reduction

As previously mentioned, we will first explore the potential gains for allowing a given accuracy loss in the compression process. We present these in table 1. Note how even an essentially lossless compression within machine precision at single precision can provide a 45% reduction in the overall data size. Note, for instance, that even allowing for a *peak* 10^{-5} error in the global dataset, can provide a 4x dataset size reduction as compared to a double precision number. Although not shown, results for the presented quantities are essentially resilient in the face of converting from double to single precision which yields an automatic 2x improvement in storage requirements and network bandwidth. On the other extreme, allowing for a fairly large, 10^{-1}, error in the whole domain requires nearly 10 times less bits than a single precision value. As we shall see, this is at the expense of some loss in accuracy but given the extreme compression ratios, the achievable results are extremely remarkable.

Table 1 Dataset sizes by allowed compression errors

Compression Accuracy (Max Error)	Dataset Size [GB]	Average Bits per Number	Savings over FP32 Number
10^{-1}	395	3.26	9.82×
10^{-2}	681	5.62	5.69×
10^{-3}	1014	8.37	3.82×
10^{-4}	1536	12.68	2.52×
10^{-5}	1945	16.06	1.99×
10^{-6}	2355	19.45	1.64×
10^{-7} (within machine precision)	2662	21.99	1.45×
Single Precision	4159	32	1.00×
Double Precision	8318	64	0.50×

B. Energy Power Spectra and Two-Point Correlations

Figures 4 & 5 highlight the effectiveness of ZFP in preserving spectral characteristics in complex datasets along the spanwise direction. Across different key regions in the boundary layer as follows: y^+ 4, 15, 250 and 500 (i.e., linear viscous layer, buffer region, log region and wake), the compression scheme is capable of preserving the one-dimensional energy power spectra of the streamwise velocity and temperature fluctuations along the spanwise direction, k_z^+, with minimal introduction of numerical artifacts. In all cases, power spectra of velocity and temperature fluctuations indicate that the turbulent flow scales have been properly resolved. The supersonic DNS case shows a significant drop off of the energy spectra, confirming that all turbulence scales were resolved. The energy-containing scale ($\sim k_z^{-1}$) and the typical -5/3 inertial range ($\sim k_z^{-5/3}$) are evidently identified. The spectral peak observed at low wavenumbers is the “footprint" signature of the outer-layer eddies or large scale motions on the near-wall region, as stated by [24]. Furthermore, it can be observed the presence of a significant zone with a local slope of $\sim k_z^{-3}$ immediately following the -5/3 inertial range or energy transfer range. According to [25], a second inertial range (i.e., enstrophy transfer range) can be found in two-dimensional turbulence, and further analysis must be carried out to physically explain this spectrum’s behavior in our three-dimensional turbulence case. This -3 power-law zone enlarges as one moves farther from the wall. The next slope corresponds to the dissipation range (-7), which is clearly resolved by DNS, as well. In addition, note the smooth “tails" of the curves at the largest waivenumbers without any sharp or abrupt end, which means that the smallest length scales of turbulence are captured (i.e., the Kolmogorov and Obukhov-Corrsin length scales for the velocity and temperature field, respectively). However, it was observed a tendency of energy spectra for all cases to pile-up in the small scale range in the very near wall region for $y^+ < 3$ (not shown). Overall, the evident similarity between E_{uu} and E_{tt} at different y^+ stations allow us to infer that Reynolds analogy is satisfied in terms of energy spectra of the velocity and temperature of the fluid flow.

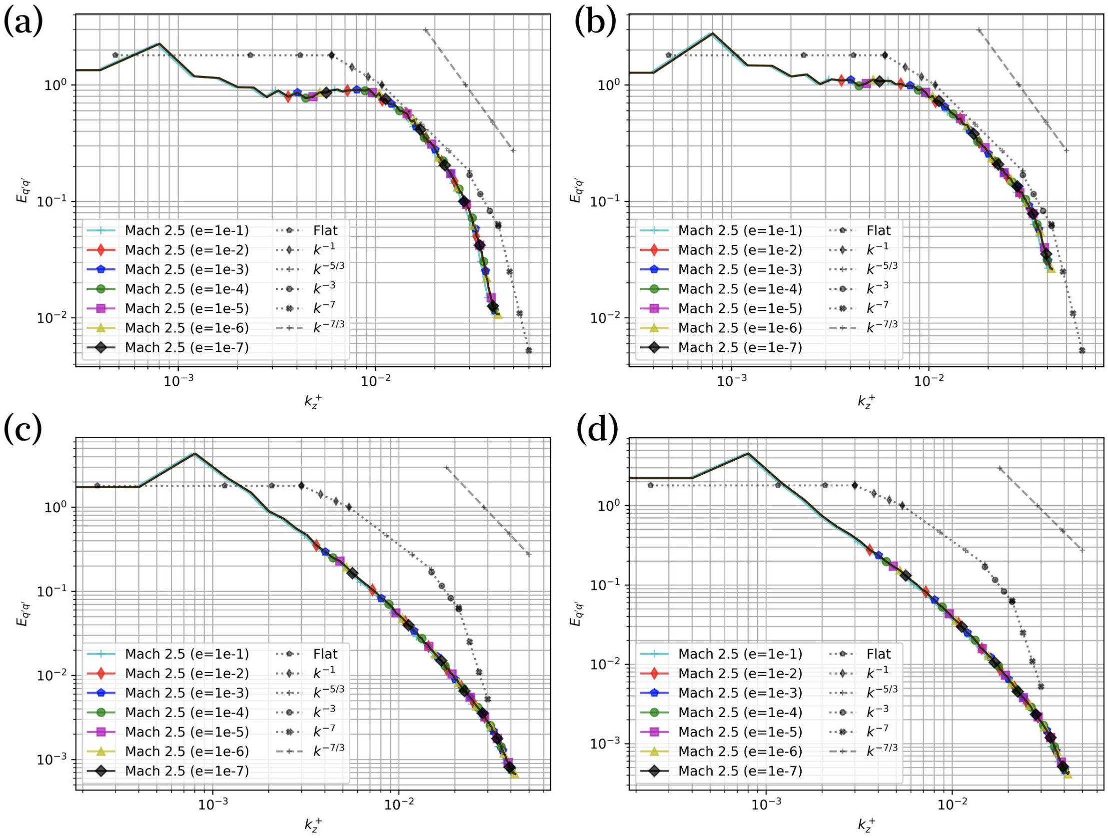


Fig. 4 Streamwise velocity fluctuation energy spectra, E_{uu}, at (a) $y^+ = 4$, (b) $y^+ = 15$, (c) $y^+ = 250$ and (d) $y^+ = 500$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

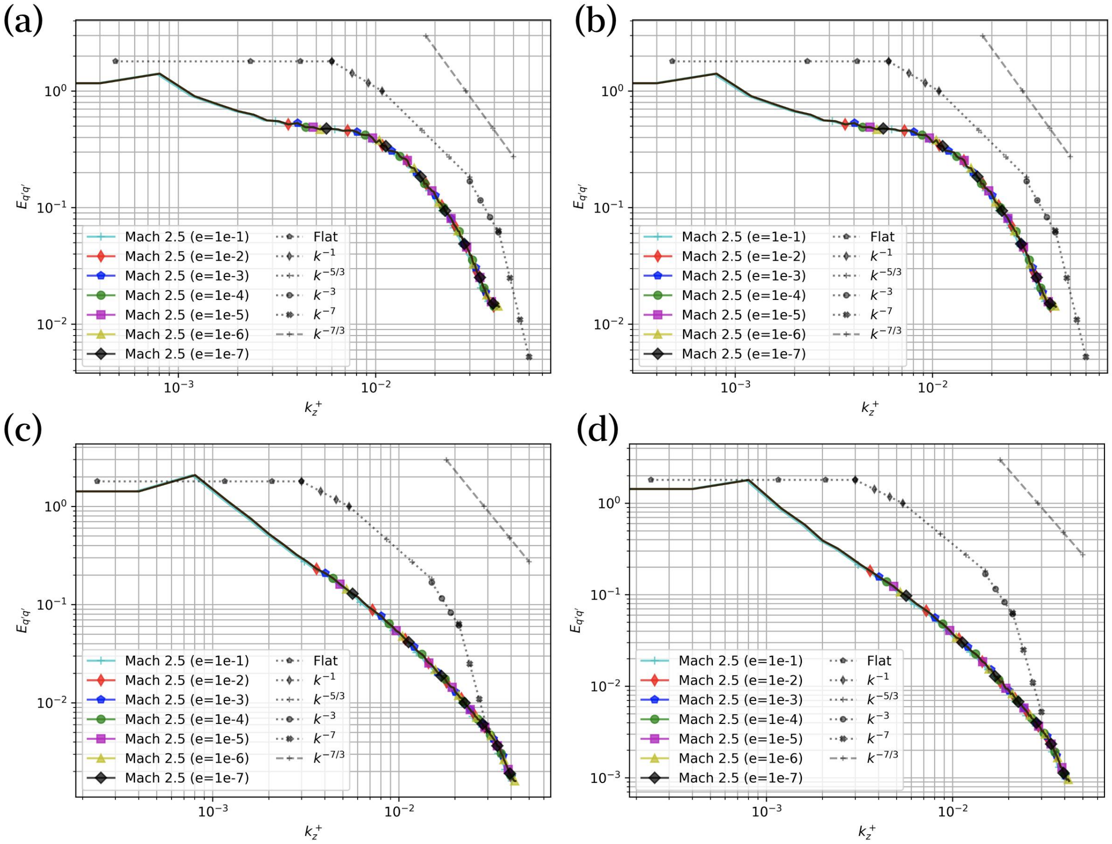


Fig. 5 Temperature fluctuation energy spectra, E_{tt}, at (a) $y^+ = 4$, (b) $y^+ = 15$, (c) $y^+ = 250$ and (d) $y^+ = 500$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

The same is true with respect to the two-point correlations shown in figures 6, 7, 8 & 9. The normal velocity fluctuation-temperature TPC, fig 10, does exhibit a small impact of the compression at very high compression ratios in the near-wall region. As we shall see in later sections, this is very likely due to a higher sensitivity in the wall-normal velocity to the compression scheme as this type of effect is somewhat amplified in parameters dependent on the normal velocity component. In all cases, it was observed that the autocorrelation and cross-correlation coefficients decay toward zero over a distance of $Lz/2$ at most (here $L_z^+ \approx 2500$), even in the outer region with the presence of wider large scale motions. This confirms that the computational domain is sufficiently wide along the spanwise direction.

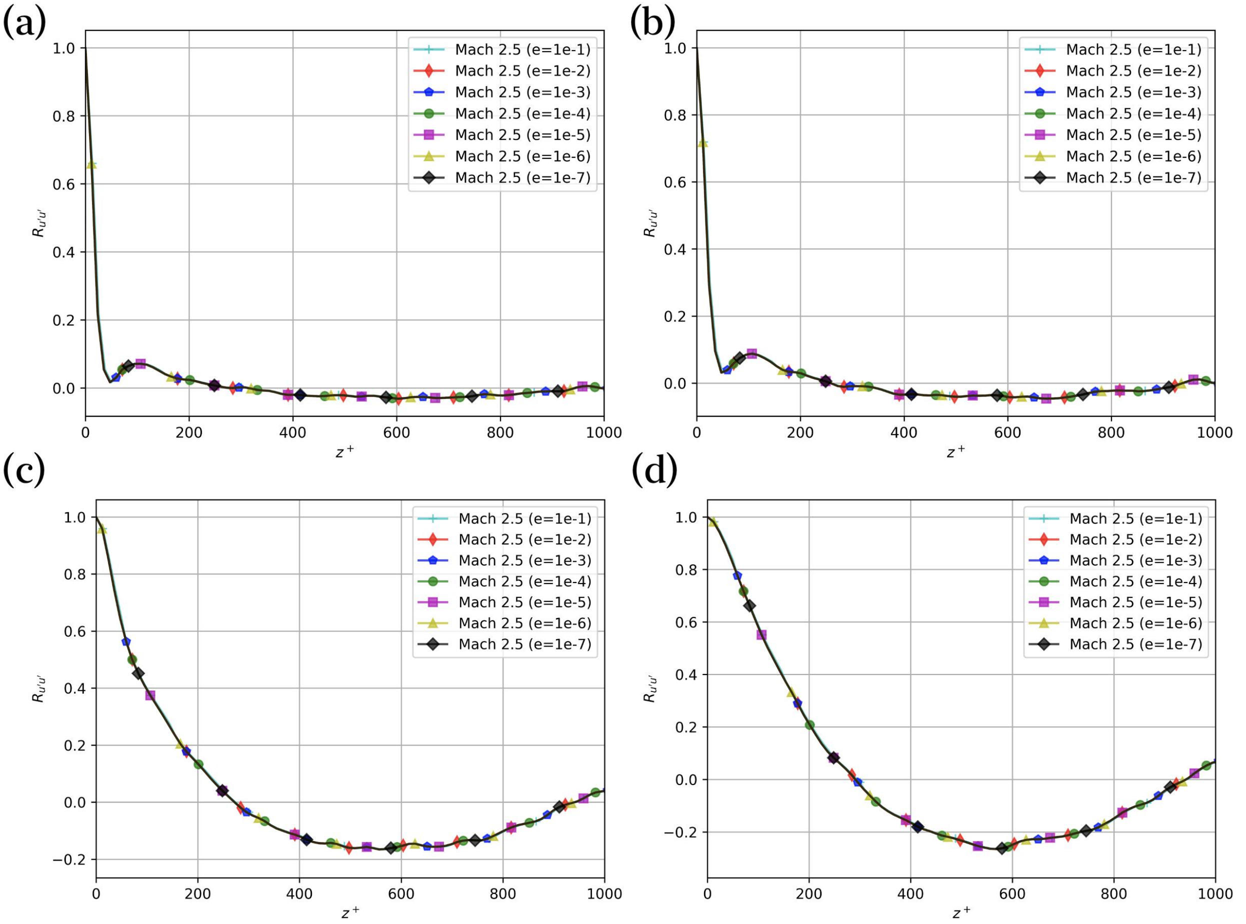


Fig. 6 Streamwise velocity fluctuation TPC at (a) $y^+ = 1$, (b) $y^+ = 15$, (c) $y^+ = 60$ and (d) $y^+ = 150$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

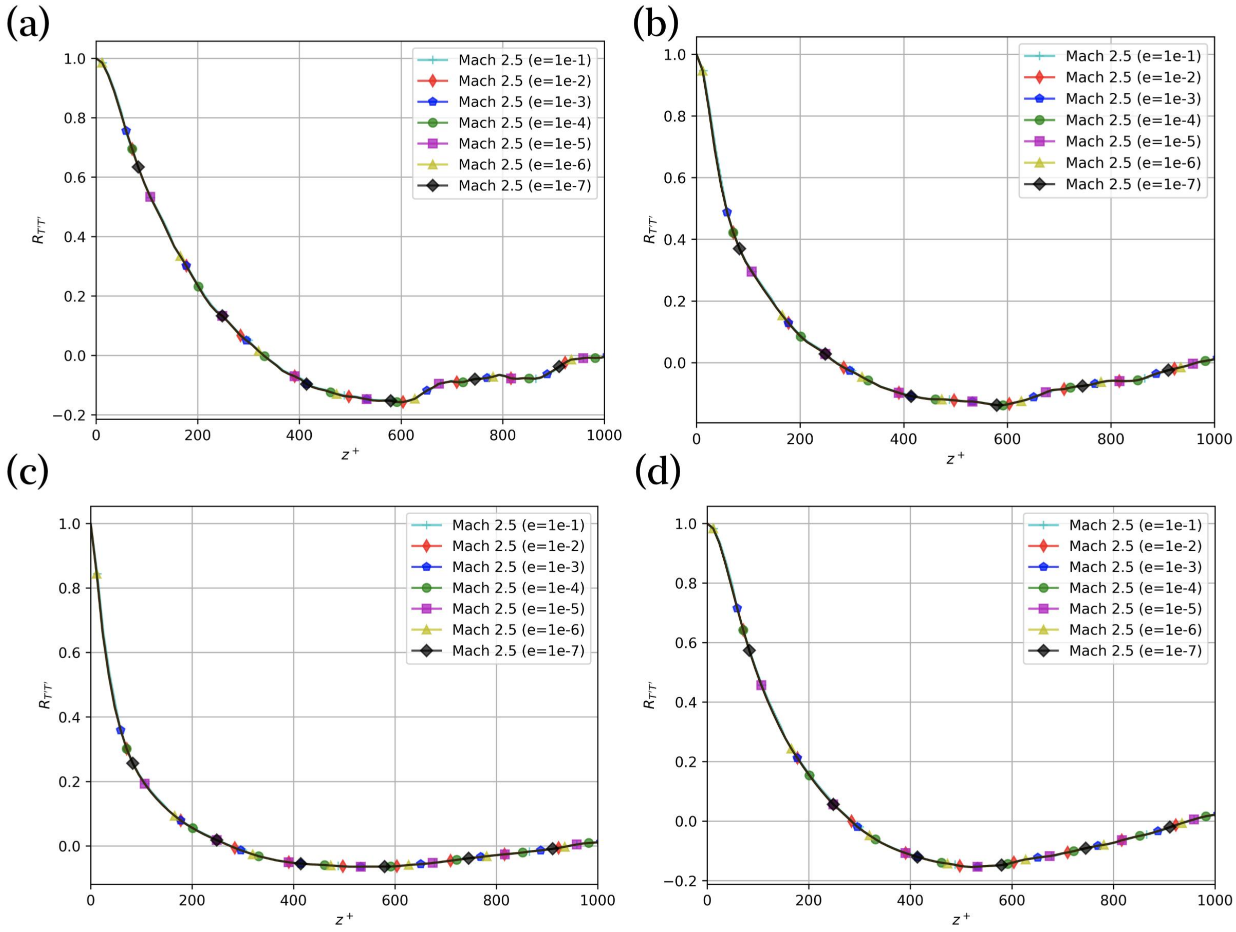


Fig. 7 Temperature fluctuation TPC at (a) $y^+ = 1$, (b) $y^+ = 15$, (c) $y^+ = 60$ and (d) $y^+ = 150$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

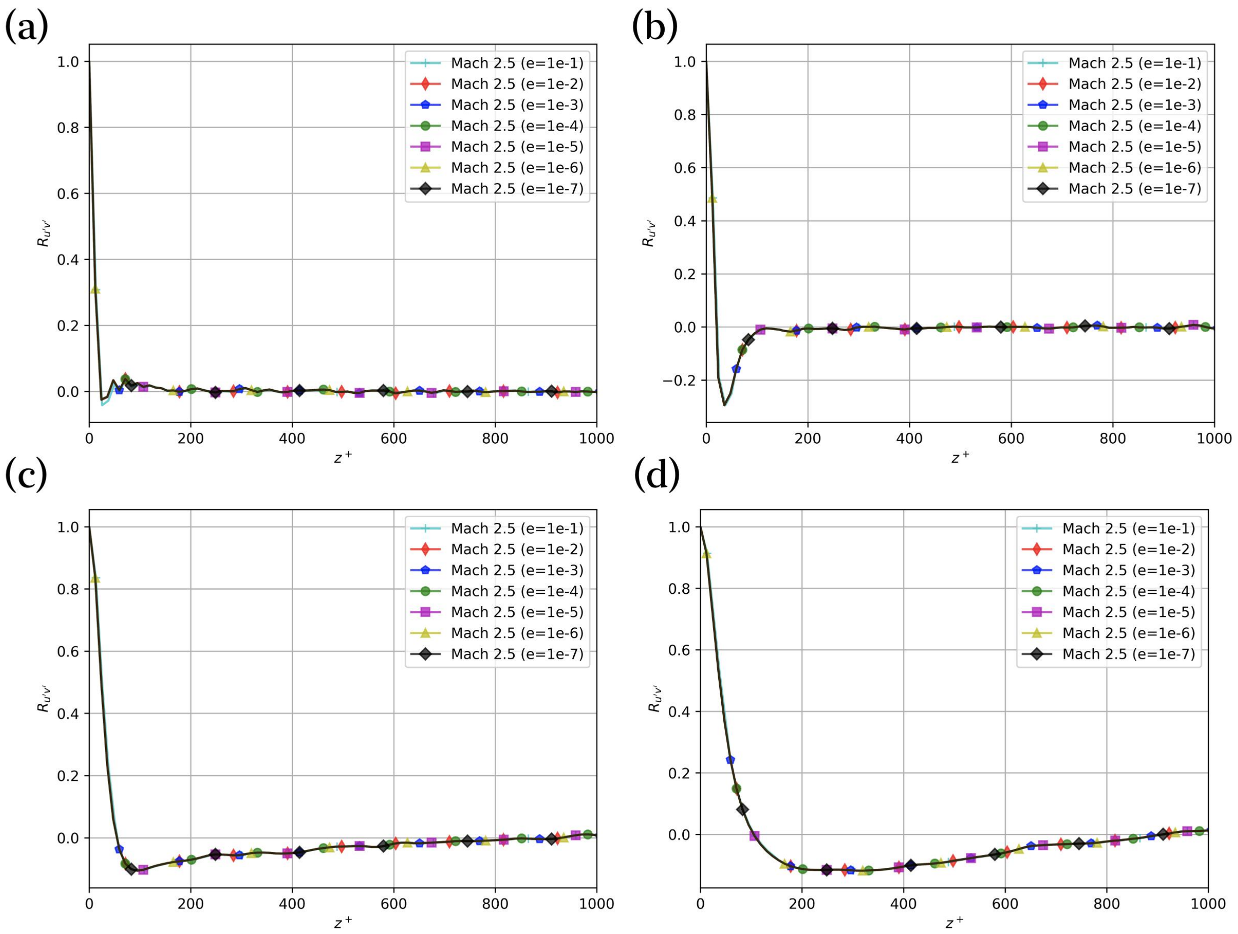


Fig. 8 Streamwise and wall-normal velocity fluctuation two-point cross-correlation at (a) $y^+ = 1$, (b) $y^+ = 15$, (c) $y^+ = 60$ and (d) $y^+ = 150$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

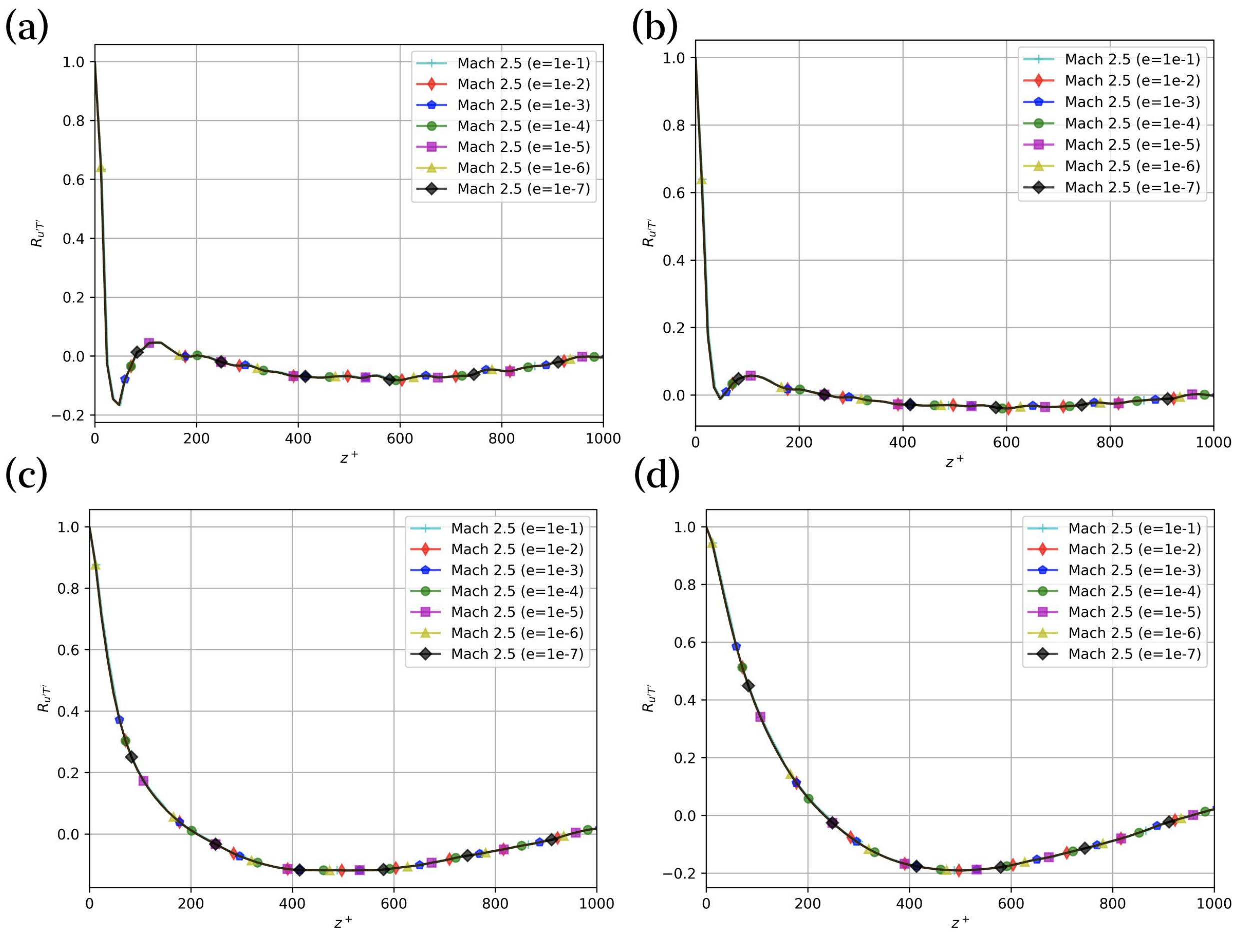


Fig. 9 Streamwise velocity and temperature fluctuation two-point cross-correlation at (a) $y^+ = 1$, (b) $y^+ = 15$, (c) $y^+ = 60$ and (d) $y^+ = 150$; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

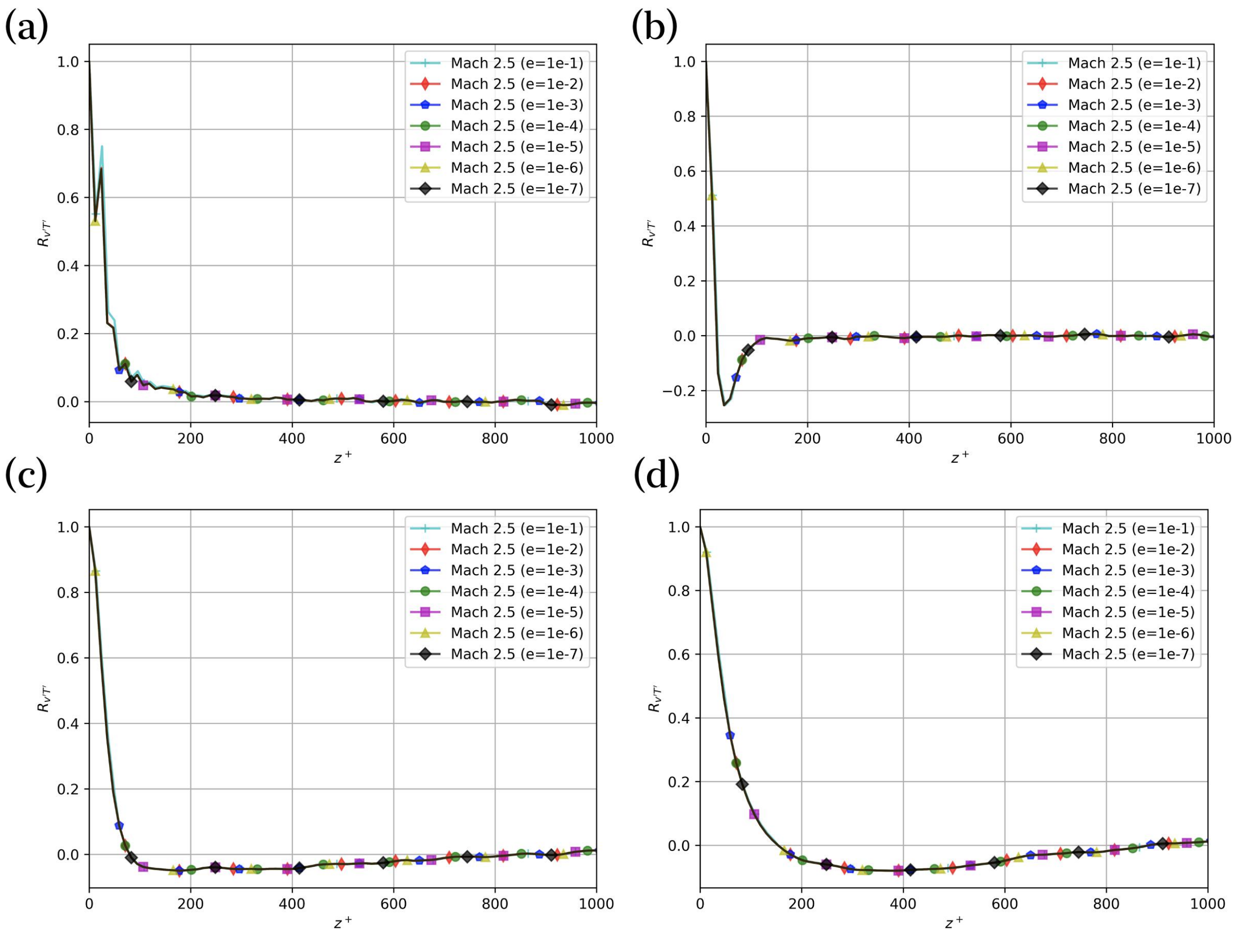


Fig. 10 Wall-normal velocity and temperature fluctuation two-point cross-correlation at (a) $y^+ = 1$**, (b)** $y^+ = 15$**, (c)** $y^+ = 60$ **and (d)** $y^+ = 150$**; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.**

C. Error Convergence for First-Order Statistics

Figures 11-15 are perhaps the most enlightening ones with respects to the practical limits of the compression scheme and its operating mechanism. Recall that in ZFP's compression scheme local values (groups of 4 per dimension) were aligned to a common exponent which can be seen as a form of quantization. This is readily observable as the values on the most extreme compression (10^{-1}) actually resemble step-like levels commonly seen in quantization. However, the power of the scheme is seen in the preservation of the overall shape and characteristics of the data per point which is achieved via local alignment rather than fitting a global, general compression transformation. This enables per-level adaptivity at the granularity of groups of 8 nodes.

We draw a parallel to the concept of convergence in numerical linear algebra to evaluate the empirical convergence rate of the ZFP error tolerance. Conceptually, one can leverage the L_p norm family to assess the convergence of the errors as compared to the baseline (within machine precision). The L_p norm is defined as,

$$L_p = ||\mathbf{x}||_p \equiv \left(\sum_i |x_i|^p \right)^{(1/p)}$$

Hence, the particular cases where $p = 1,\ 2\ \&\ \infty$ result in

$$L_1 = ||\mathbf{x}||_1 \equiv \sum_i |x_i|$$

$$L_2 = ||\mathbf{x}||_2 \equiv \sqrt{\sum_i |x_i|^2}$$

$$L_\infty = ||\mathbf{x}||_\infty \equiv \lim_{p\to\infty}\left(\sum_i |x_i|^p\right)^{(1/p)} = \max(x)$$

where **x** represents the error of a given flow variable of interest measured against the machine precision value at 10^{-7}. In figs. 11-15, we further normalize the norms and present the inner-scaled norms. It is worth noting that the L_p norms are representative of the total error. The mean error is roughly two orders of magnitude smaller than those shown in the figures below. Normalizing the difference with the local value at machine precision (not shown) results in a similar trend for the norms.

It is important to highlight the norm used in determining a given convergence rate since there are no formal guarantees connecting rates in one norm to that of a different one. We show graphical results for L_2 and L_∞ norms in figs. 11-15 and a summarized result for L_1 norm in table 2.

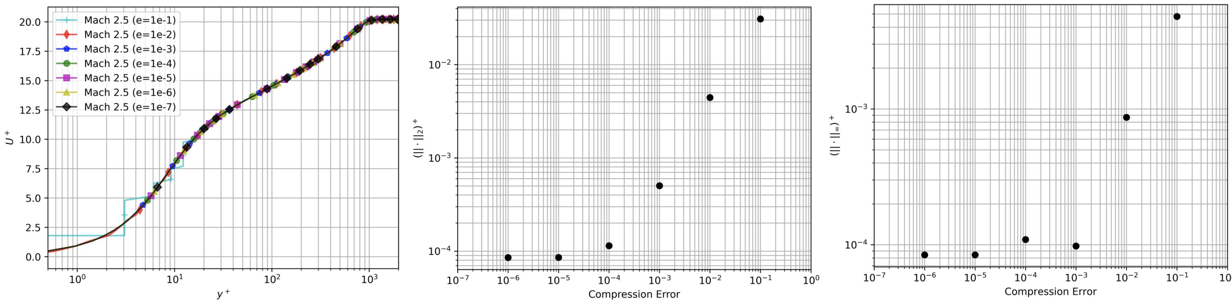


Fig. 11 (Left) Inner-scaled velocity; (center) L_2 Convergence; (right) L_∞ Convergence; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

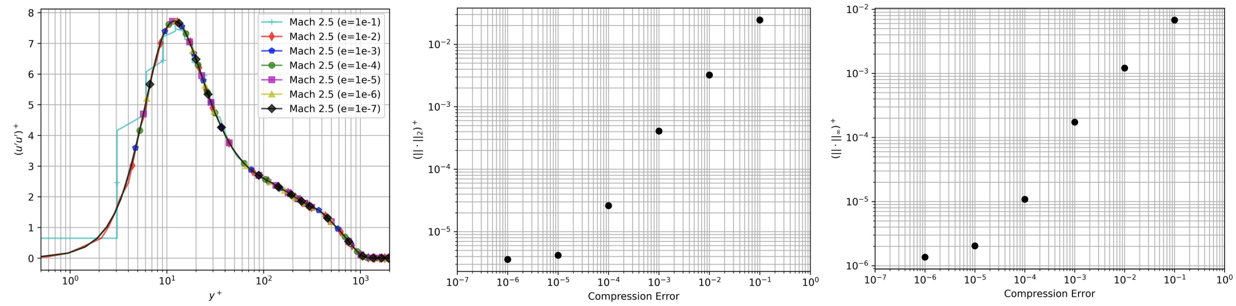


Fig. 12 (Left) Inner-scaled $u'u'$; (center) L_2 Convergence; (right) L_∞ Convergence; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

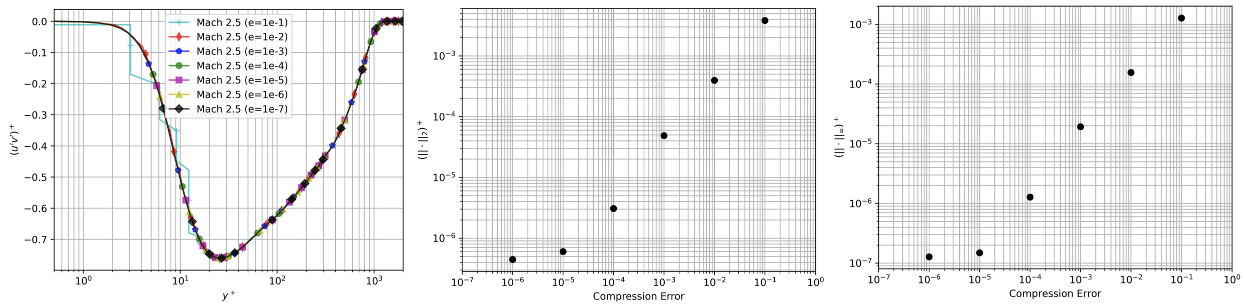


Fig. 13 (Left) Inner-scaled $u'v'$; (center) L_2 Convergence; (right) L_∞ Convergence; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

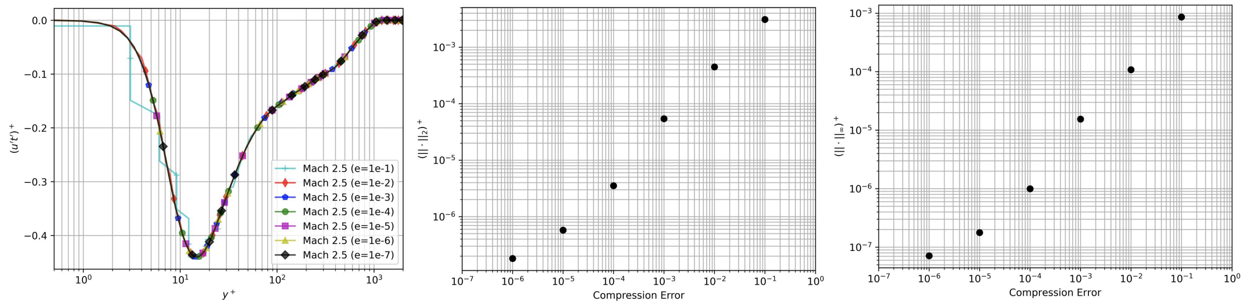


Fig. 14 (Left) Inner-scaled $u'T'$; (center) L_2 Convergence; (right) L_∞ Convergence; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

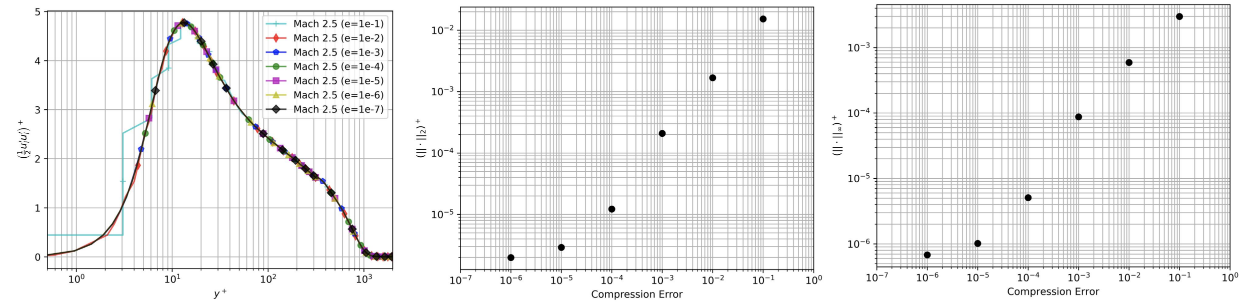


Fig. 15 (Left) Inner-scaled turbulent kinetic energy; (center) L_2 Convergence; (right) L_∞ Convergence; Note: Symbols are added to aid visibility and do not represent the actual resolution of the sampled data points.

The results suggest the practical limit for adimensionalized CFD results to be 10^{-2} for the Reynolds/Mach number pairing considered. In general, we observed the error convergence leveled off at 10^{-5} suggesting this might be a reasonable trade-off for practical compression of data for long term archival. As discussed in table 1, this threshold would yield a 4× reduction in storage requirement over traditional double precision floating point data. This would translate to storing 4 times more flow fields or requiring a quarter of the storage. It is also worth noting that this is nearing the storage requirements of a half-precision (FP16) floating point number without the shortcomings of such a standard format due to ZFP's ability to dynamically compress floats with spatial locality and with an emphasis on derived quantities such as gradients of compressed datasets.

Although we do not delve into a theoretical analysis of the convergence in different norms, we do present an empirical study of the error convergence in different norms in table 2. In the L_1, L_2 and L_∞ norms, the method is best described as a second-order compression scheme with perhaps the only outlier being the inner-scaled velocity. It is worth highlighting that the largest errors are seen in the near-wall region where the strongest gradients are observed. These large gradients likely highlight the limitations in the compression scheme since lossy compression of numbers with similar magnitudes is a far simpler challenge compared to regions with strong variations. Figures 11-15 clearly highlights this limitation with the appearance of somewhat piecewise averaging (or stepping) for the highest compression ratio case. On average, the tendency and magnitudes are preserved allowing for a general picture of the BL behavior with a dataset almost 20× smaller than a traditional FP64 dataset (storing nearly 3 bits per float on average).

Table 2 Convergence rates on L_1 (not shown in plots), L_2 and L_∞

Variable	L_1 Convergence Rate	L_2 Convergence Rate	L_∞ Convergence Rate
U^+	1.96	1.70	1.34
$u'u'$	2.16	2.05	1.88
$u'v'$	2.26	2.16	2.12
$u'T'$	2.02	2.06	2.06
TKE	2.26	2.13	1.83

D. Higher-Order Statistics

Up to this point, we have only considered first and second order data statistics. Higher order statistics provide additional information on the distribution of the data considered in a statistical manner. To this end, we consider skewness (third-order) and flatness or kurtosis (fourth-order) as two higher order statistics. We present the results in figures 16-20. We again see the remarkable performance of ZFP and its resilience in preserving data statistics and qualities from first to higher-order statistics. For both skewness and kurtosis, the larger compression ratios do exhibit more notable errors. In particular, the wall-normal velocity, fig 18, exhibits very notable oscillations towards the outer region of the boundary layer and also drastic discrepancies at the wall.

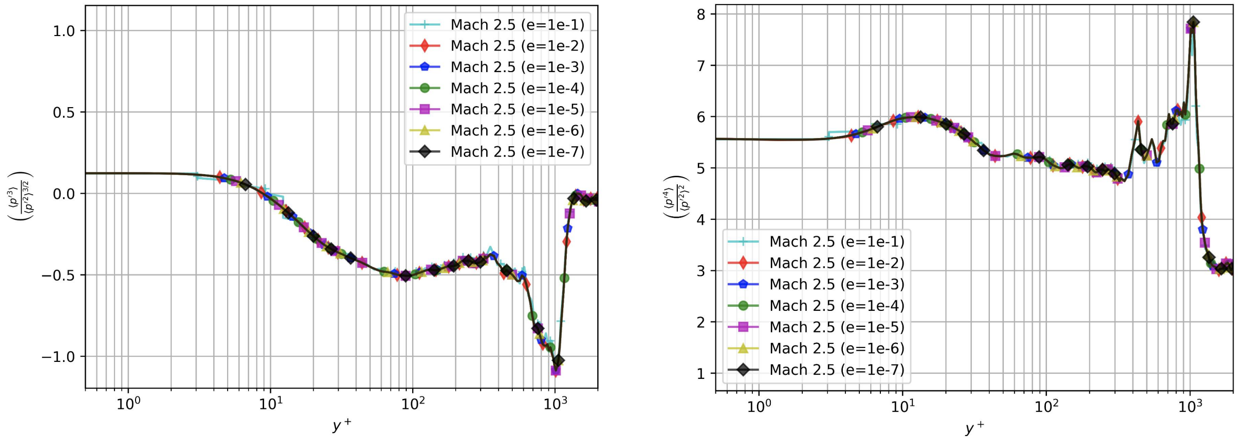


Fig. 16 (Left) Inner-scaled skewness (left) and flatness (right) for the pressure fluctuation

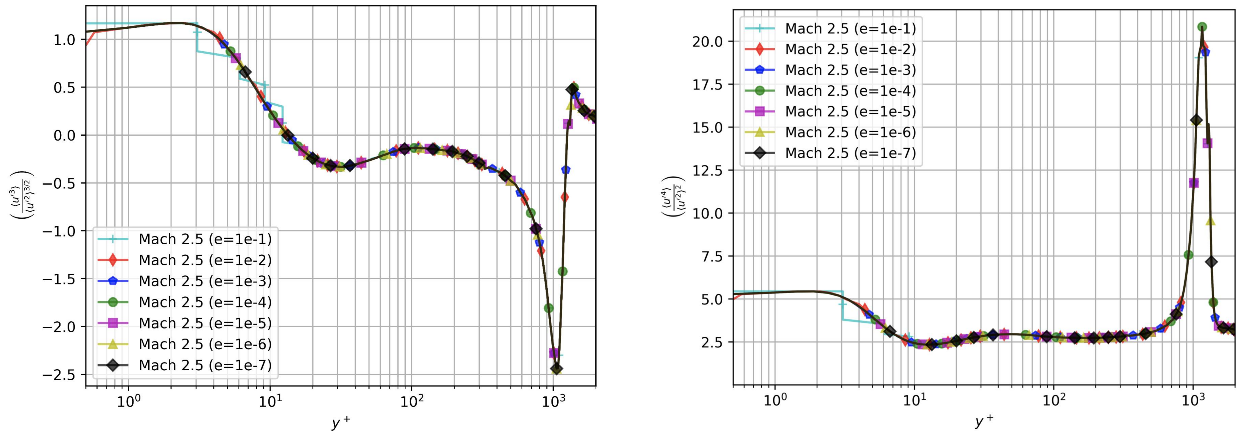


Fig. 17 (Left) Inner-scaled skewness (left) and flatness (right) for the streamwise velocity fluctuation

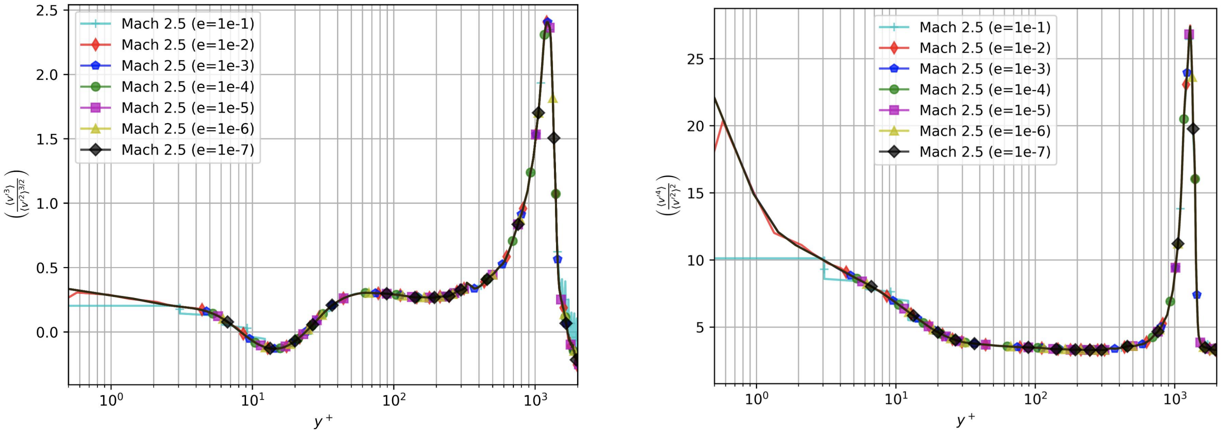


Fig. 18 (Left) Inner-scaled skewness (left) and flatness (right) for the wall-normal velocity fluctuation

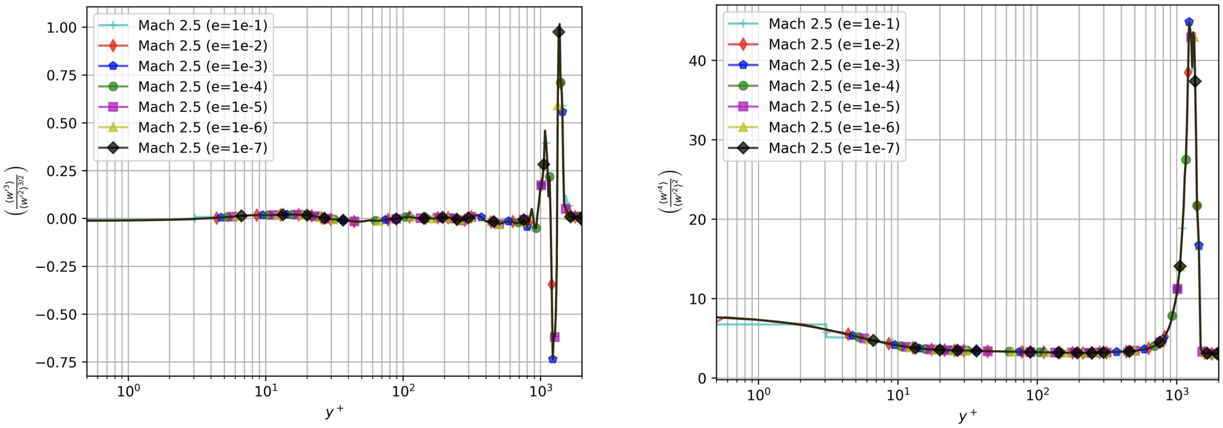


Fig. 19 (Left) Inner-scaled skewness (left) and flatness (right) for the spanwise velocity fluctuation

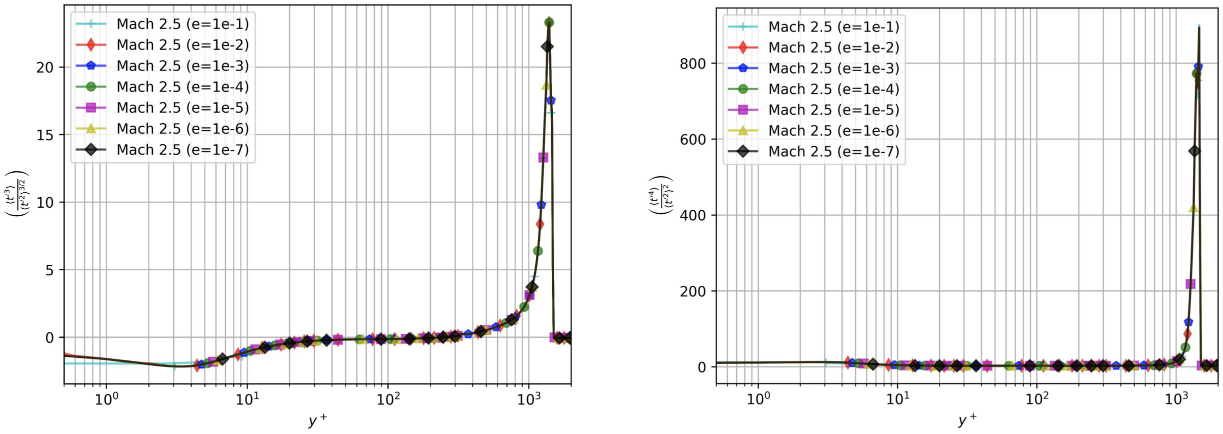


Fig. 20 (Left) Inner-scaled skewness (left) and flatness (right) for the temperature fluctuation

For these higher-order statistics (figs. 16-20), the qualitative tolerance is higher than that of first-order statistics. Beyond 10^{-3}, including more precision bits results in diminishing returns. The valuable takeaway lesson is that the target compression should be very dependent on the precision and quality requirements. General trends are preserved and a first-order inspection might benefit greatly from high compression ratios such as the ones attained with the 10^{-2} or 10^{-3} targets. Looking back at table 1, this represents 7.64x to 11.38x less storage than the accustomed FP64 (double precision) format often demanded in scientific applications with diminishing returns in many post-processed quantities. Bare in mind that the 10^{-3} target is in the realm of roughly one byte per float on average.

V. Conclusion

We have evaluated the performance of a locally adaptive lossy compression scheme for floating point data, ZFP, on a Mach 2.5, spatially developing turbulent boundary layer at a moderately high Reynolds number. As compression schemes have evolved, the ability to apply lossy compression to DNS data and achieve high fidelity results is now within reach and is readily available via plugins to standard computing tools such as HDF5. Interestingly, we found point-wise parameters calculated in the physical domain were more prone to compression errors. Nonetheless, the overall trend of the data is preserved tho quantization phenomena are evidently present at very high compression ratios. Spectral quantities are surprisingly well preserved even at very high compression ratios; something not achievable via downsampling-based schemes. Two-point correlations are also remarkably resilient in the face of large admissible errors during compression. We have also highlighted that higher order statistics are also well-preserved. In summary, we have evaluated a general, floating-point compression scheme on a high-fidelity numerical dataset. The results suggest that the scientific community could benefit from considering and evaluating ZFP as a potential lossy compression scheme to include for long term archival or increased network performance at the expense of minimal accuracy loss once the tolerable accuracy losses are prescribed.

Acknowledgments

Authors acknowledge resources from award CAWT-IRG3 (NSF #1849243). Christian Lagares acknowledges financial support from the National Science Foundation under grant no. HRD-1906130. Guillermo Araya acknowledges financial support from the National Science Foundation CAREER award no. 1847241 and AFOSR grant no. FA9550-22-1-0089. This work was supported in part by a grant from the DoD High-Performance Computing Modernization Program (HPCMP).

References

[1] Araya, G., and Lagares, C., "Implicit subgrid-scale modeling of a Mach-2.5 spatially-developing turbulent boundary layer," *Entropy*, Vol. 24, No. 555, 2022. URL `https://doi.org/10.3390/e24040555`.

[2] ISO/IEC 10918-1:1994, "Information technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines," Standard, International Organization for Standardization, Geneva, CH, Feb. 1994.

[3] Mark Adler *et al.*, "Portable Network Graphics (PNG) Specification (Second Edition)," Specification, W3C, November 2003. Https://www.w3.org/TR/2003/REC-PNG-20031110/.

[4] Wilson, J. P., "Lossy Compression of Turbulence Data," , 1998.

[5] Otero, E., Vinuesa, R., Marin, O., Laure, E., and Schlatter, P., "Lossy Data Compression Effects on Wall-bounded Turbulence: Bounds on Data Reduction," *Flow, Turbulence and Combustion*, Vol. 101, No. 2, 2018. https://doi.org/10.1007/s10494-018-9923-5.

[6] Lindstrom, P., "Fixed-Rate Compressed Floating-Point Arrays," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, 2014. https://doi.org/10.1109/TVCG.2014.2346458.

[7] Hutchins, N., and Marusic, I., "Evidence of very long meandering features in the logarithmic region of turbulent boundary layers," *Journal of Fluid Mechanics*, Vol. 579, 2007, pp. 1–28.

[8] Araya, G., Castillo, L., Meneveau, C., and Jansen, K., " A dynamic multi-scale approach for turbulent inflow boundary conditions in spatially evolving flows," *Journal of Fluid Mechanics*, Vol. 670, 2011, pp. 518–605.

[9] Araya, G., Lagares, C., and Jansen, K., "Direct simulation of a Mach-5 turbulent spatially-developing boundary layer," *49^{th} AIAA Fluid Dynamics Conference, AIAA AVIATION Forum, (AIAA 3131876) 17 - 21 June, Dallas, TX*, 2019.

[10] Lagares, C., and Araya, G., "Compressibility Effects on High-Reynolds Coherent Structures via Two-Point Correlations," *AIAA AVIATION Forum, (AIAA-3516309) 2 - 6 August (Virtual)*, 2021.

[11] Lund, T., Wu, X., and Squires, K., "Generation of turbulent inflow data for spatially-developing boundary layer simulations," *Journal of Computational Physics*, Vol. 140, No. 2, 1998, pp. 233–258.

[12] Lagares, C. J., and Araya, G., "Power spectrum analysis in supersonic/hypersonic turbulent boundary layers," *AIAA SCITECH 2022 Forum*, 2022. https://doi.org/10.2514/6.2022-0479.

[13] G. Urbin and D. Knight, "Large-Eddy Simulation of a supersonic boundary layer using an unstructured grid," *AIAA Journal*, Vol. 39, No. 7, 2001, pp. 1288–1295.

[14] Stolz, S., and Adams, N., "Large-eddy simulation of high-Reynolds-number supersonic boundary layers using the approximate deconvolution model and a rescaling and recycling technique," *Physics of Fluids*, Vol. 15, No. 8, 2003, pp. 2398–2412.

[15] Whiting, C. H., Jansen, K. E., and Dey, S., "Hierarchical basis in stabilized finite element methods for compressible flows," *Comp. Meth. Appl. Mech. Engng.*, Vol. 192, No. 47-48, 2003, pp. 5167–5185.

[16] Jansen, K. E., "A stabilized finite element method for computing turbulence," *Comp. Meth. Appl. Mech. Engng.*, Vol. 174, 1999, pp. 299–317.

[17] Araya, G., Castillo, C., and Hussain, F., "The log behaviour of the Reynolds shear stress in accelerating turbulent boundary layers," *Journal of Fluid Mechanics*, Vol. 775, 2015, pp. 189–200.

[18] Doosttalab, A., Araya, G., Newman, J., Adrian, R., Jansen, K., and Castillo, L., "Effect of small roughness elements on thermal statistics of a turbulent boundary layer at moderate Reynolds number," *Journal of Fluid Mechanics*, Vol. 787, 2015, pp. 84–115.

[19] Araya, G., Lagares, C. J., Santiago, J., and Jansen, K. E., "Wall temperature effect on hypersonic turbulent boundary layers via DNS," *AIAA Scitech 2021 Forum*, 2021. https://doi.org/10.2514/6.2021-1745.

[20] Araya, G., Lagares, C., and Jansen, K., "Reynolds number dependency in supersonic spatially-developing turbulent boundary layers," *2020 AIAA SciTech Forum (AIAA 3247313) 6 - 10 January, Orlando, FL.*, 2020.

[21] Lagares, C., Rivera, W., and Araya, G., "Scalable Post-Processing of Large-Scale Numerical Simulations of Turbulent Fluid Flows," *Symmetry*, Vol. 14, No. 4, 2022. URL https://www.mdpi.com/2073-8994/14/4/823.

[22] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," , 2015. URL https://www.tensorflow.org/, software available from tensorflow.org.

[23] The HDF Group, “Hierarchical data format version 5,” , 2000-2010. URL http://www.hdfgroup.org/HDF5.

[24] Pirozzoli, S., and Bernardini, M., “Turbulence in supersonic boundary layers at moderate Reynolds number,” *Journal of Fluid Mechanics*, Vol. 688, 2011, pp. 120–168.

[25] Kraichnan, R. H., “Inertial-range transfer in two- and three-dimensional turbulence,” *Journal of Fluid Mechanics*, Vol. 47, No. 3, 1971, p. 525–535. https://doi.org/10.1017/S0022112071001216.