

# TOPIO

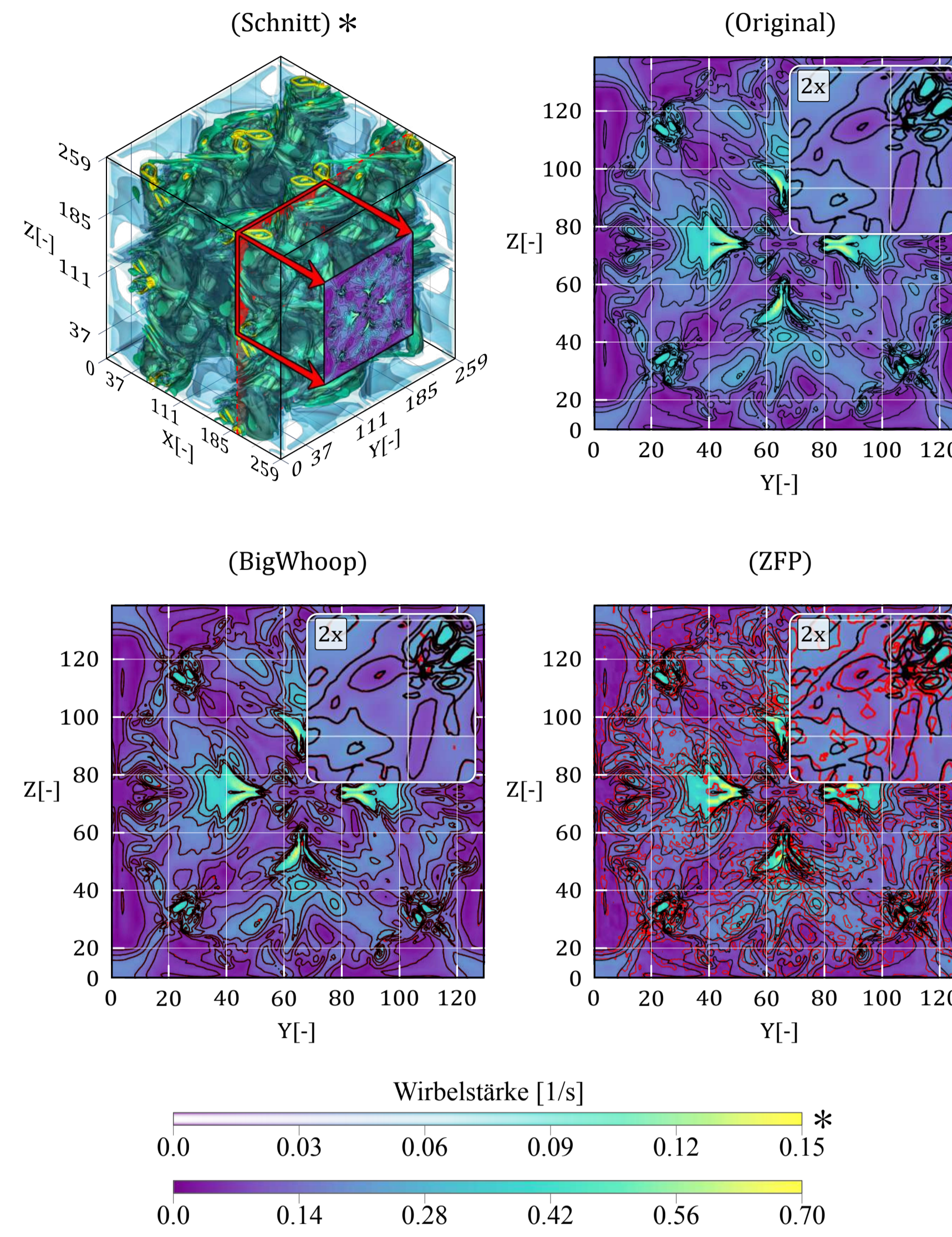
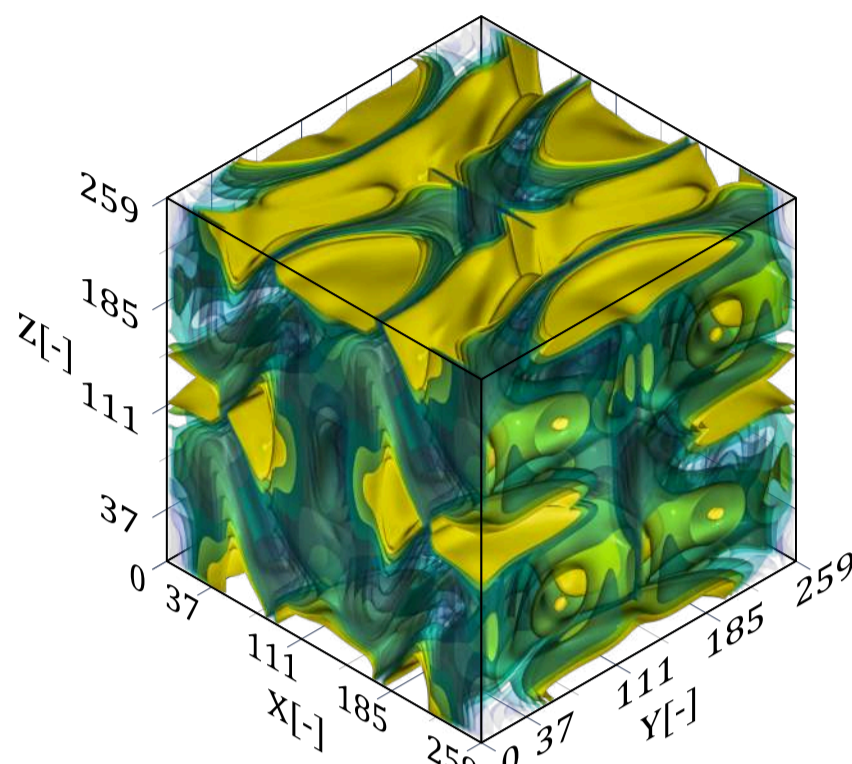
## Analyse der I/O-Optimierungs- und Kompressionsstrategien

### KOMPRESSSION

#### Grundlage der Analyse

Der Taylor-Green Wirbel (TGV) stellt ein einfaches und wohl definiertes hydrodynamisches Problem dar, dessen breites Spektrum turbulenter Skalen eine Analyse der Auswirkungen der Längenskala auf die Leistung der einzelnen Kompressoren ermöglicht.

- Analytische Ausgangslösung auf einem  $256^3$  Zellen großen Gitter.
- Die Auswertung erfolgt für die dimensionslosen Zeitschritte  $t = 0, 25, 50, 75, 100, 125, 150, 175$  und  $200$ .
- Größe pro Zeitschritt: 703.052.304 Bytes.



#### Wirbelstärke

Konturdiagramme der Wirbelstärke im Vergleich zwischen dem ursprünglichen und dem komprimierten Datensatz für den dimensionslosen Zeitschritt  $t = 75$  des TGV-Testfalls. Die Position des Schnitts ist in der isometrischen Ansicht (oben links) angegeben. Die Komprimierung wurde mit einem Verhältnis von 100:1 unter Verwendung der BWC- und ZFP-Bibliotheken durchgeführt. Konturlinien werden verwendet, um den Unterschied zwischen dem ursprünglichen (schwarz) und dem komprimierten (rot) Feld anzudeuten.

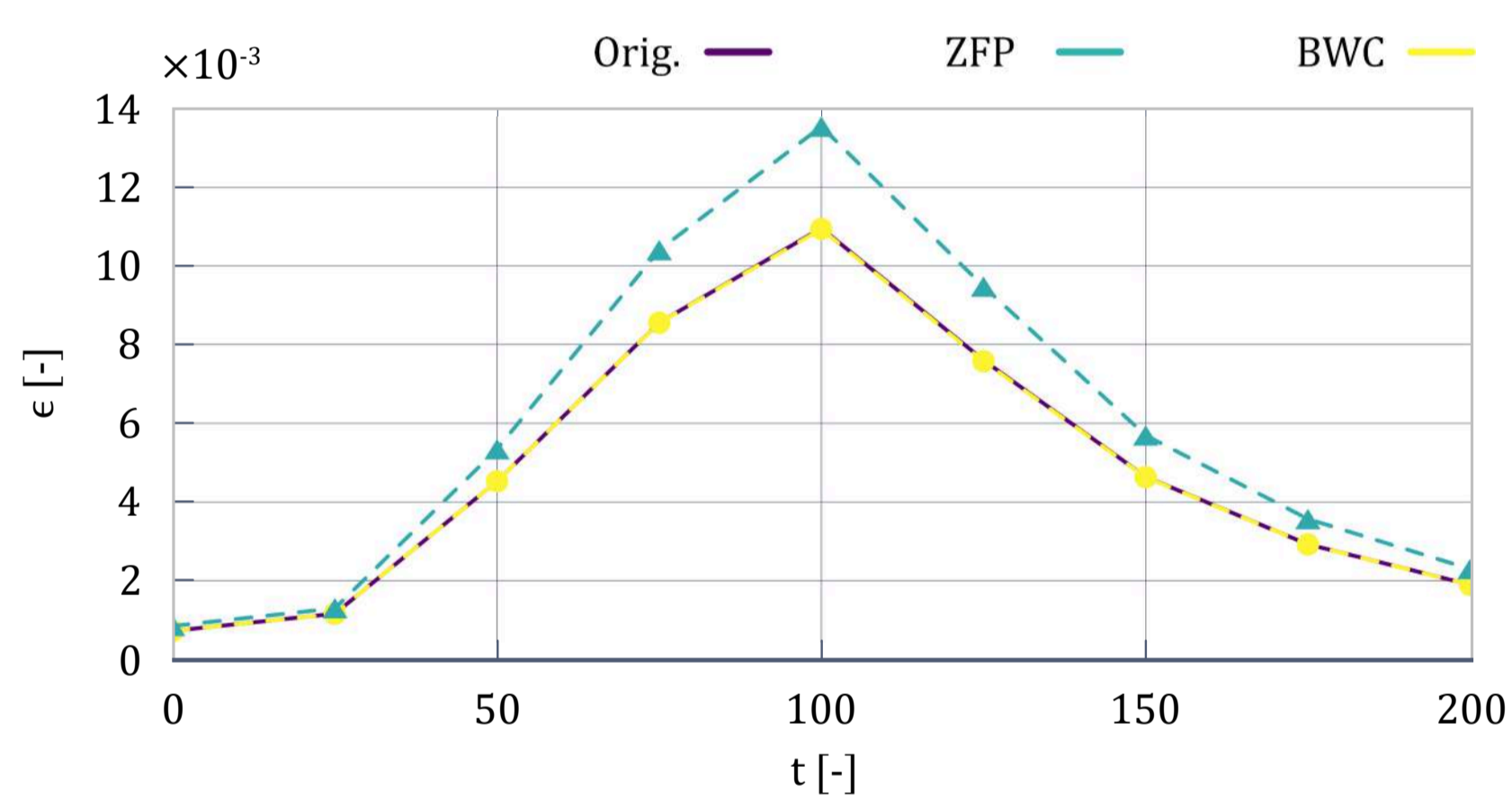
Hervorzuheben ist das die BigWhoop Bibliothek auch für abgeleitete Größen wie die Wirbelstärke noch eine gute Rekonstruktion ermöglicht. Sowohl eine visuelle als auch eine analytische Auswertung ist somit auch für hohe Kompressionsverhältnisse möglich.

Die ZFP-Bibliothek zeigt dagegen starke Abweichungen, die selbst eine visuelle Beurteilung der Wirbelstrukturen nicht mehr zulassen.

Kompressionsverhältnis, Kompressionszeit und die Peak Signal-to-Noise Ratio (PSNR) für die Kompressoren BigWhoop, ZFP und 7-Zip. Die besten Ergebnisse sind grün dargestellt.

Zeitschritt	Kompressionsverhältnis			Kompressionszeit			Peak Signal-to-Noise Ratio		
	BWC	ZFP	7-Zip	BWC	ZFP	7-Zip	BWC	ZFP	7-Zip <sup>a</sup>
0	64.7	65.0	41.7	6.9	1.7	114.9	133.9	61.2	∞
25	62.2	62.1	6.0	4.4	2.0	160.4	123.6	60.0	∞
50	61.6	61.1	5.8	7.2	2.0	173.2	92.0	48.2	∞
75	61.0	61.1	5.6	8.8	1.9	164.0	73.8	44.3	∞
100	60.8	61.1	5.1	11.0	1.8	172.9	62.6	37.9	∞
125	60.8	61.1	4.6	11.4	1.8	172.8	63.3	40.6	∞
150	60.8	61.1	4.2	11.1	1.8	186.4	64.1	40.5	∞
175	60.8	61.1	3.9	10.7	1.8	183.9	65.7	43.0	∞
200	60.9	61.1	3.7	11.1	1.9	181.5	67.1	45.2	∞

<sup>a</sup> Das PSNR für den 7-Zip-Kompressor ist aufgrund seines verlustfreien Charakters immer unendlich.



#### Energiedissipation

Dargestellt ist die normalisierte Energiedissipationsrate als Funktion der Wirbelentwicklung des Taylor-Green Wirbels über die dimensionslose Zeit  $t$ . Der originale Datensatz wurde mit BWC- und ZFP-komprimierten Datensätzen bei einem Kompressionsverhältnis von 120:1 verglichen.

Die Energiedissipationsrate gibt die Energiemenge an, die durch die viskosen Kräfte in einer turbulenten Strömung dissipiert wird. Sie gibt uns also ein Maß dafür, wie viel Energie in den kleinsten turbulenten Strukturen konzentriert ist. Wie erwartet, zeigt sich in den Vergleichsstudien in der obigen Tabelle, dass Zeitschritte mit einer hohen Energiedissipationsrate einer schlechteren Verdichtungsleistung entsprechen. Außerdem müssen wir feststellen, dass die BWC-Bibliothek die ursprüngliche Energiedissipationsrate recht gut wiedergeben kann, während die Anwendung der ZFP-Bibliothek zu einer Überschätzung führt.

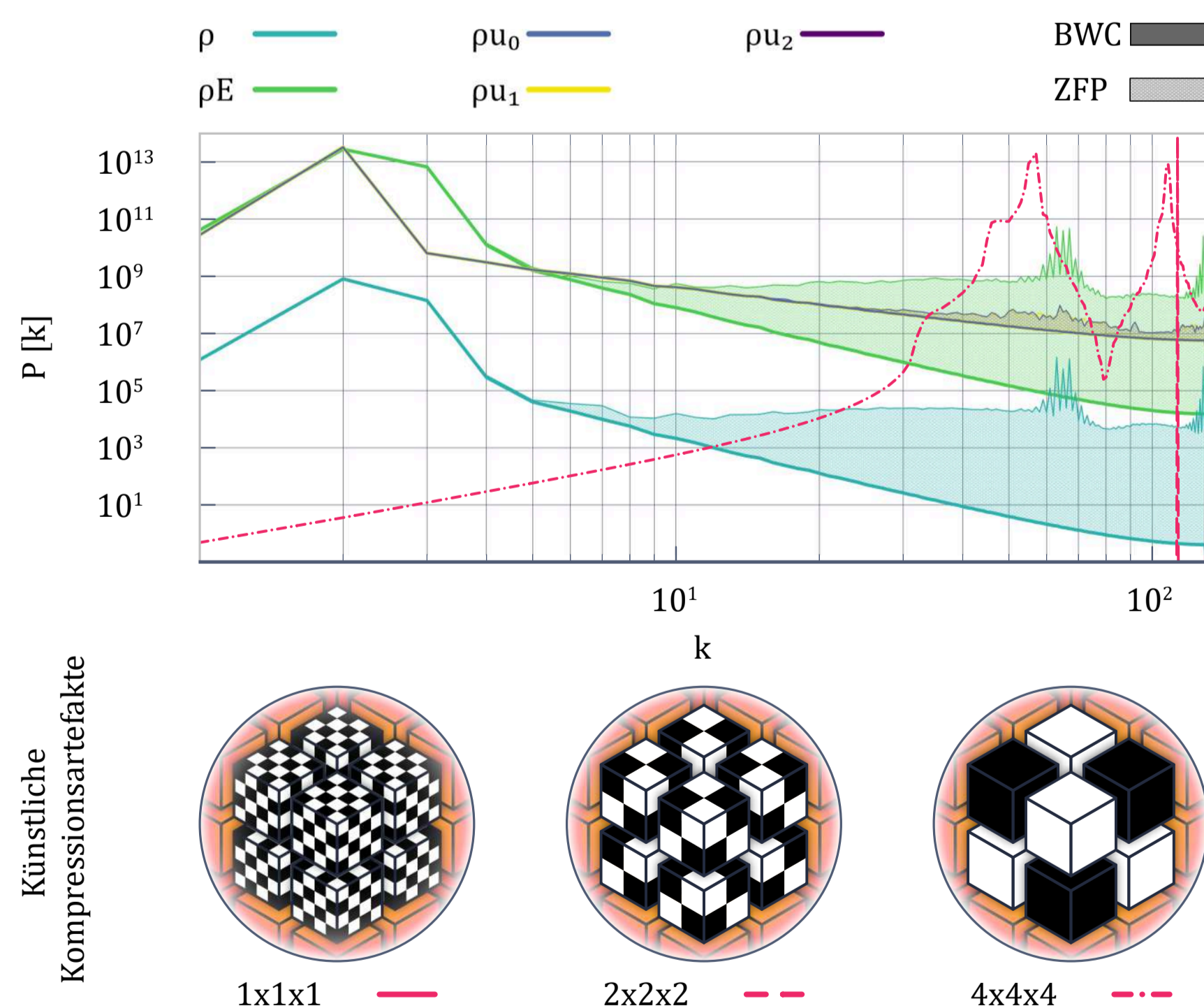
turbulenten Strukturen konzentriert ist. Wie erwartet, zeigt sich in den Vergleichsstudien in der obigen Tabelle, dass Zeitschritte mit einer hohen Energiedissipationsrate einer schlechteren Verdichtungsleistung entsprechen. Außerdem müssen wir feststellen, dass die BWC-Bibliothek die ursprüngliche Energiedissipationsrate recht gut wiedergeben kann, während die Anwendung der ZFP-Bibliothek zu einer Überschätzung führt.

#### Leistungsspektrum

Um die Überschätzung der Energiedissipationsrate einordnen zu können wurde ein Leistungsspektrum für den dimensionslosen Zeitschritt  $t = 0$  des Taylor-Green-Wirbel-Testfalls erstellt. Der ursprüngliche Datensatz wurde mit BWC- und ZFP-komprimierten Datensätzen mit einem Kompressionsverhältnis von 120:1 verglichen. Des Weiteren wurde das Leistungsspektrum von künstlichen Kompressionsartefakten, mit einer Größe von  $1 \times 1 \times 1$ ,  $2 \times 2 \times 2$  und  $4 \times 4 \times 4$  Datenpunkten, ausgewertet, die typisch für blocktransformationsbasierte Codecs (z.B. ZFP) sind.

Aus dem Diagram lässt sich ableiten, dass die ZFP Bibliothek in den Bereichen eine signifikant höhere Leistung vorhersagt, die für eine hohe Kompressionsrate von Kompressionsartefakten dominiert sind.

Eine Auswertung des visuellen Einflusses der Kompressionsartefakte erfolgt anhand der Wirbelstärke.



### I/O OPTIMIERUNG

#### I/O-Softwarestack

Die Optimierung der I/O-Aktivitäten einer Anwendung und eine effiziente parallele Datenübertragung sind anspruchsvolle Aufgaben. Es gibt komplexe Abhängigkeiten zwischen den Schichten des I/O-Softwarestacks. Jede Ebene bietet mehrere Konfigurationsparameter. Die Konfiguration dieser Parameter hängt von verschiedenen Faktoren ab. Diese sind z. B. Data Striping oder I/O-Muster. Nutzer können ihre Anwendungen möglicherweise nicht optimal abstimmen. Wenn Standardeinstellungen verwendet werden, führt dies häufig zu einer schlechten I/O-Effizienz.

Im Anwendungsfall MPAS<sup>1</sup> gibt es viele Konfigurationsparameter:

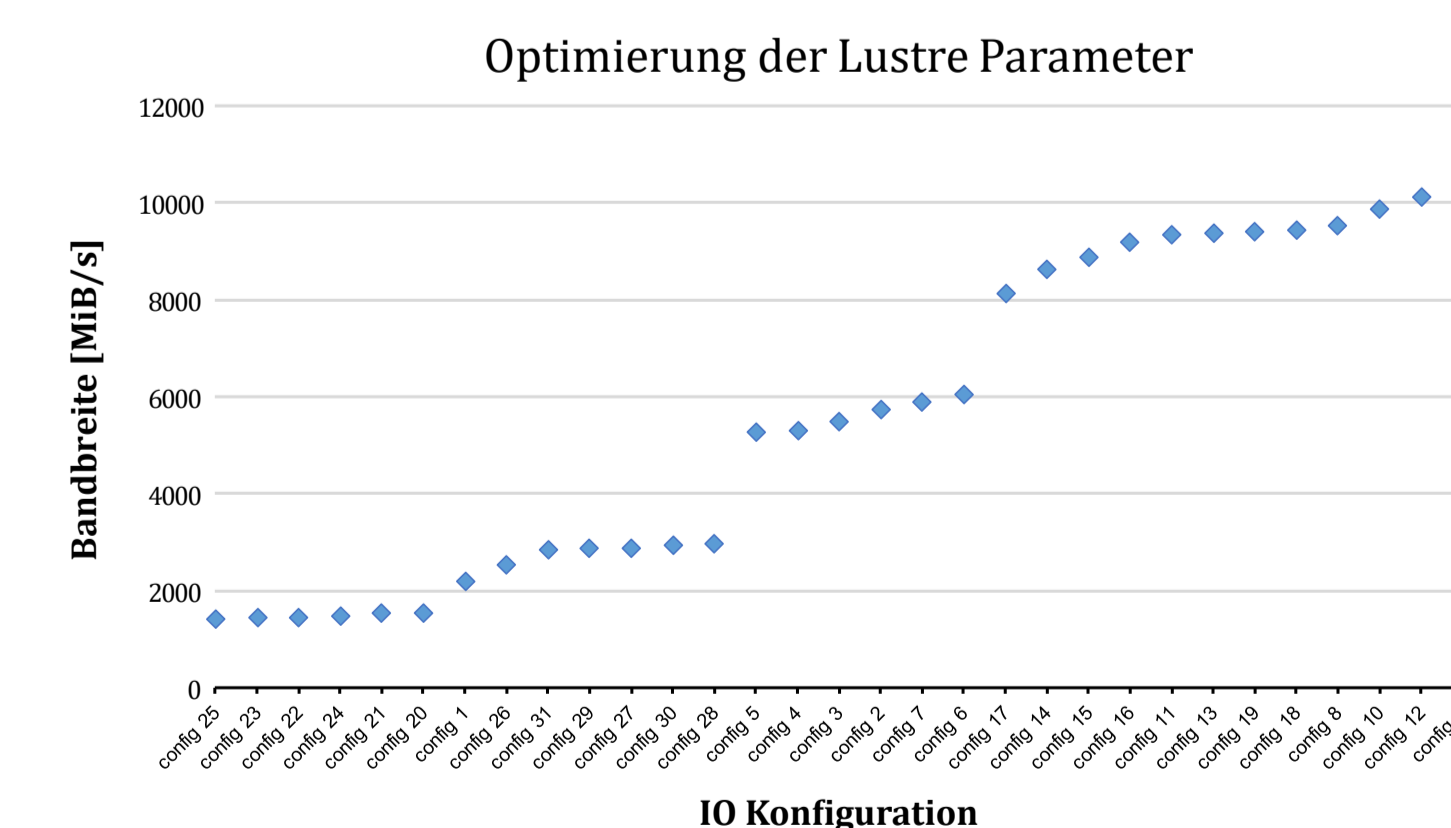
- In der Anwendung wird die Anzahl an I/O-Prozessen und deren Verteilung festgelegt.
- Für die I/O Bibliothek pNetCDF müssen Größe und Anordnung von Datenpaketen gewählt werden.
- Im I/O-Backend-Interface MPI-IO können Collective Buffering und Data Sieving aktiviert oder deaktiviert werden. Für Collective Buffering wird die Größe des Buffer-Speichers und die Anzahl der Buffer-Knoten gewählt.
- Beim parallelen Dateisystem ist das Data Striping von Bedeutung. Dabei werden Dateien im Dateisystem in Segmenten abgelegt. Es ermöglicht paralleles Lesen und Schreiben auf mehreren Laufwerken. Der Parameter Stripe Count gibt die Anzahl der Hardware Ressourcen an, auf welche die Dateien verteilt werden. Mit Stripe Size wird die Größe der Datensegmente in Bytes definiert.

[1] <https://github.com/MPAS-Dev/MPAS-Model>

#### Manuelle Optimierung des I/O: kürzere Laufzeit

Von den oben genannten Konfigurationsparametern wurde für eine erste manuelle Optimierung das Data Striping variiert. Als Standardeinstellung auf dem Lustre File-System WS11 des HLRS ist das progressive File-Layout gewählt. Hierfür wurde eine Laufzeit von 222,3 Sekunden gemessen. Mit der Optimierung konnte die Gesamtlaufzeit für den Testfall auf 61% reduziert werden ohne Veränderungen am Code vorzunehmen.

Stripe Anzahl	Gesamtlaufzeit [s]					
	Stripe Größe [MB]					
	1	2	4	8	16	32
4	245,0	237,1	241,8	242,6	236,6	243,6
8	184,4	177,4	178,2	178,4	180,0	177,1
16	150,4	149,8	148,3	147,8	149,9	150,1
32	138,2	136,7	134,2	136,0	136,0	135,5
40	139,5	135,8	137,1	142,6	137,0	139,3



#### Manuelle Optimierung des I/O: höhere Bandbreite

Die Bandbreite des Schreibvorgangs wurde mit dem Profiling-Tool Darshan gemessen. Die Schreibrate der 3D Ausgabedatei mit einer Größe von 26 GB konnte um den Faktor 4,7 erhöht werden. Es wurde eine Bandbreite von 10 325.94 MiB/s erreicht.

#### Autotuning

Für die Optimierung der Striping Parameter war eine Testreihe mit 31 Tests nötig. Für eine Optimierung mit allen I/O-Parametern erreicht der Testumfang enorme Ausmaße. Mit dem Autotuning Ansatz soll die Prognose der I/O-Performance möglich sein, sodass nicht alle Kombinationen getestet werden müssen.

#### Projektkoordinator

Höchstleistungsrechenzentrum Stuttgart (HLRS)  
 Prof. Dr. Michael M. Resch  
 Nobelstraße 19  
 D-70569 Stuttgart  
 Tel: +49 711 685-87200  
 E-Mail: resch@hlrs.de

#### Projektvolumen

0.75 Mio. €  
 BMBF-Förderung: 0.75 Mio. € (100 %) mit zus. 0,15 Mio. € an sogenanntem Projektkostenzuschuss..  
 Gefördert im Rahmen des Förderschwerpunkts "Neue Methoden und Technologien für das Exascale-Höchstleistungsrechnen" (SCALEXA).

#### Projektdauer

01.11.2022 bis 31.10.2025

#### Projektpartner

- Höchstleistungsrechenzentrum Stuttgart (HLRS), Universität Stuttgart
- Institut für Physik und Meteorologie (IPM), Universität Hohenheim

#### Kontakt

Anna Mack, Patrick Vogler - HLRS  
 E-Mail: anna.mack@hlrs.de, patrick.vogler@hlrs.de  
 Arun Kumar Dwivedi - IPM  
 E-Mail: arunkumar.dwivedi@uni-hohenheim.de