

# 乱流大規模直接数値計算コードの SX-Aurora TSUBASA上での性能評価

山梨大学大学院総合研究部  
山本 義暢

概要：

世界最大Re条件での壁乱流DNSデータベースの構築を目指し、  
低メモリ量化及び超並列計算が可能なDNSコード開発の実施  
(開発要件は、世界最大・最速DNSコードとなる)

# 1. 研究背景と目標

## ➤ 研究背景

右図に示す通り、高レイノルズ数( $Re$ )壁面乱流場における乱流輸送現象の解明は学術・工学的に重要であり、第一原理的手法である乱流の**直接数値計算**(Direct Numerical Simulation, **DNS**)が最も有力な手法である。

しかし近年のスーパーコンピュータは乱流のDNSに適したシステムとは言えず  
(**メモリ規模・通信性能**)、**Re規模の拡大が困難**になっている



## ➤ 研究目標

**世界最大 $Re_\tau=16000$** , さらにはその数倍規模のDNSデータベース構築が可能な**高速・大容量DNSコードの開発**

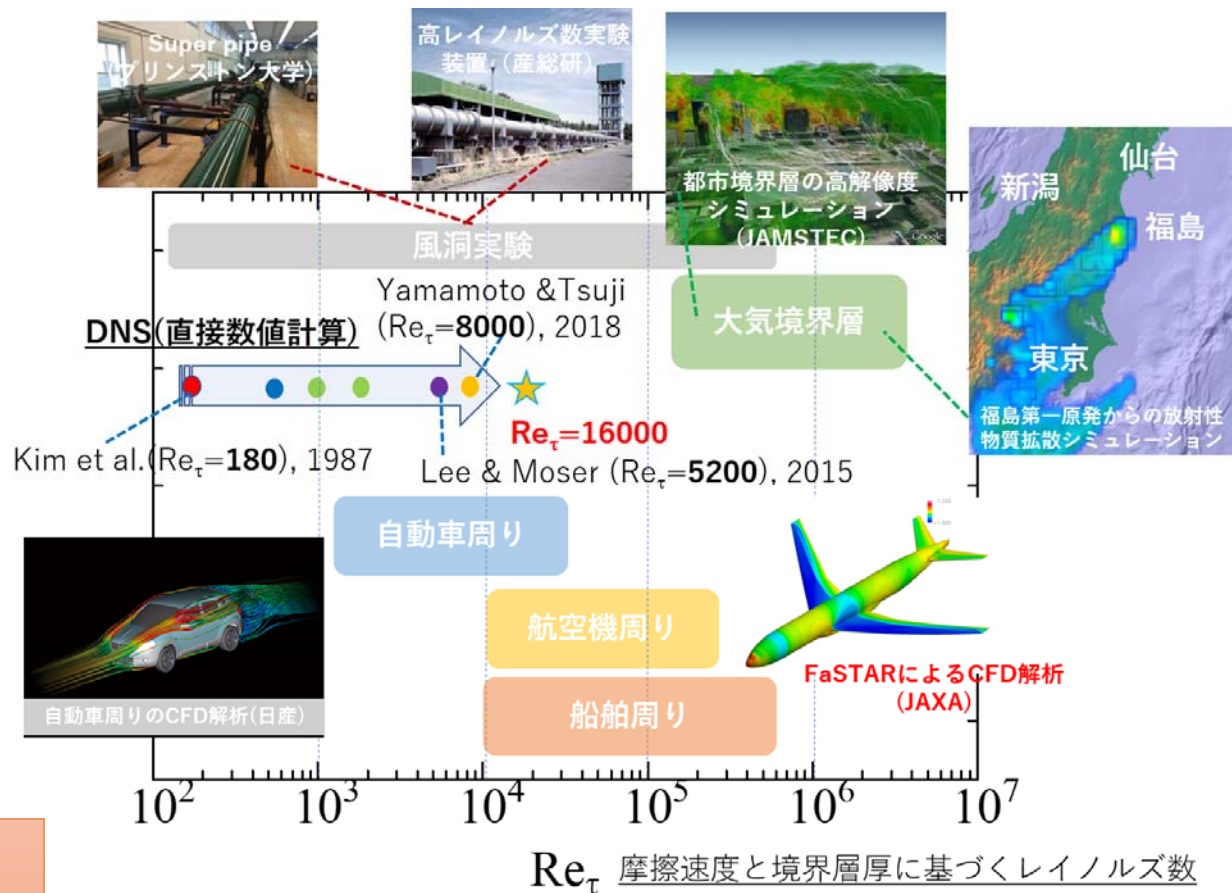


図 壁乱流及び各種工学流れにおける $Re$ 規模  
Deck et al. J. Fluid Mech. (2014) を基に作成

## 2. 高レイノルズ数DNSの問題点（現状）

表  $Re_\tau=16000$ 規模のDNSの計算コスト(演算量・メモリ量)

$Re_\tau$	$(x, y)$	$z$	aliasing error	$N_x$	$N_y$	$N_z$	Tflop /sub-step	メモリ量 [TB]
8000	Fourier	2 <sup>nd</sup> FDM	3/2-rule	6912	5760	4096	216	60
16000	Fourier	2 <sup>nd</sup> FDM	3/2-rule	13824	11520	5760	<b>1215</b>	<b>340</b>

表 全国共同利用施設における計算機性能: TOP500, 2021/6より

	システム	#node or #VE	#cores	メモリ量 [TB]	理論性能 [Tflop/s]	HPCG [Tflop/s]	HPCG 効率[%]
東大	FX1000	7,680	368,640	240	22,121	818	3.7
東大	CX1640	8,178	556,104	<b>767</b>	13,555	529	2.8
名大	FX1000	2,304	110,592	72	6,618	386	3.5
東北大	SX-ACE	1,024	4,096	64	262	28	10.6
東北大	次期システム	?	?	?	20,000	?	?
海洋研	SX-AT	5,472	43,776	257	9,991	748	7.5

### ➤ $Re_\tau=16000$ のDNSにおける問題点

- ・全系を使用しても**メモリ不足**（そもそもシステム更新でメモリが全く増えない）
- ・実アプリ(HPCG)での**低い実行効率**

### 3. DNSコード開発における課題：開発要件

Re<sub>t</sub>=16000規模のDNSを全国共同利用施設を用いて実施するには以下の対応が不可欠

#### [1] 低容量化

- ・要件：メモリ規模（200TB）で動作するための低容量化  
340TB→200TBへのメモリ量低減(現状から約4割減)

#### [2] 超並列計算への対応

- ・要件：50万コア以上での並列計算が可能  
2次元領域分割コード → スペクトル法では**all-to-all通信コスト対策**が必要

#### [3] 高速演算

- ・要件：長時間時間積分を実行可能とするための、高速演算  
1200[Tflop/step] 3[s/step] → **400[Tflop/s] 級の実行演算速度**
- ・参考：チャンネル流における最速コードは**180[Tflop/s]** Lee et al. SC2013

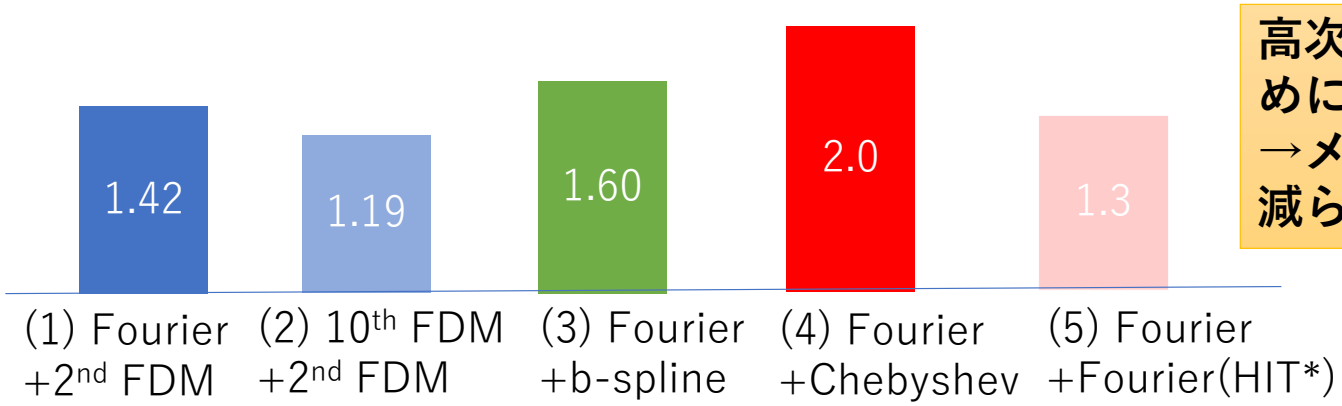
本研究では、現状コード分析を行ったうえで、上記3要件を満たす次世代DNSコード開発を実施する

# 4. 現状分析(1) 計算コスト

表 計算手法別によるDNSの演算コスト(チャンネル流 実測値)

	$Re_\tau$	$(x, y)$	$z$	aliasing error	$N_x$	$N_y$	$N_z$	$N_{xyz} \times 10^9$	Tflop /sub-step
(1)	8000	Fourier	2 <sup>nd</sup> FDM	3/2-rule	6912	5760	4096	163	216
(2)	8000	10 <sup>th</sup> FDM	2 <sup>nd</sup> FDM	/	8640	6144	4096	217	240
(3)	5200	Fourier	b-spline		3/2-rule	10240	7680	1536	121
(4)	2560	Fourier	Chebyshev	3/2-rule	1024	1024	1024	1	2.0
(5)	HIT*	Fourier	Fourier	PS	1024	1024	1024	1	1.3

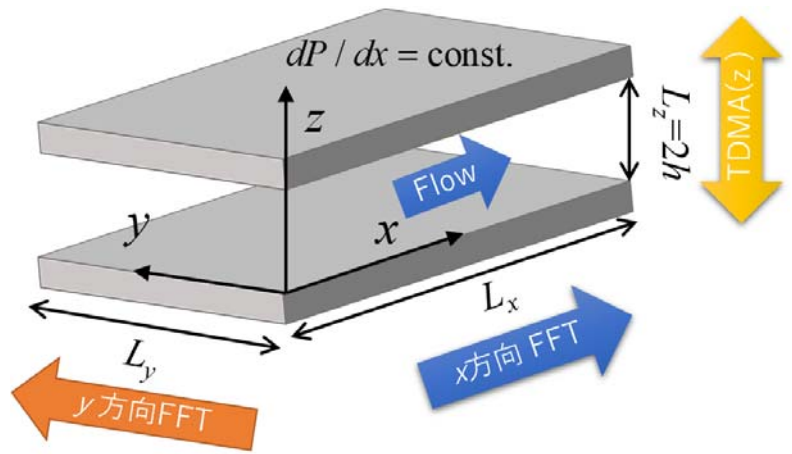
(1) Kaneda & Yamamoto, J. Fluid Mech. submitted, (2) Yamamoto & Tsuji, Phys. Rev. Fluids(2018),  
 (3) Lee & Moser, J. Fluid Mech.(2015), (4). Morishita et al., J. Phys. Soc. Jpn(2019),  
 (5) Yokokawa et al. SC02(2002), \*Homogeneous Isotropic Turbulence



高次精度差分では同一精度を確保するために格子数が増える  
 →メモリ量増加、演算コストもそれほど減らない(**Fourier + 2<sup>nd</sup> FDMが有力**)

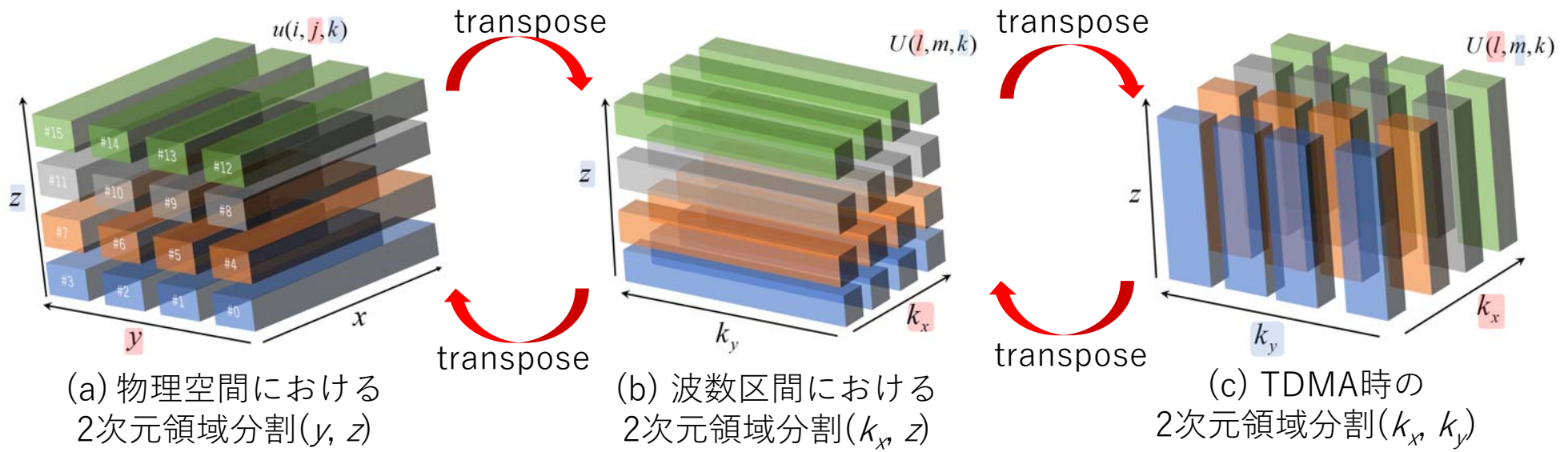
FIG. 計算手法別の1024<sup>3</sup> 格子換算 1 sub-step当たりの計算コスト (Tflop/sub-step)

# 4. 現状分析(2) 通信コスト



**問題点**：スペクトル法ベースのDNSコードに2次元領域分割を適用すると下図のような**2段階の軸転置**が発生し、通信コストが極めて高い  
 →近年のスパコンは演算性能の向上に比べ、通信性能の向上も極めて鈍い

FIG. チャンネル流DNSにおける主要演算部



## 5. 開発状況 (1)低容量化

(1)スペクトル法におけるaliasing誤差除去方法変更による低容量手法の開発

(2)非線形項計算順序の再考によるメモリ量低減化

(3)非線形項2重計算によるメモリ量低減化  
→但し演算量は増える

(4)統計処理コードのポスト処理化によるメモリ量低減化

(5)通信・演算バッファ細分化によるメモリ量低減化



**約4割程度のメモリ量低減化に成功**

**Re<sub>τ</sub>=16000規模のDNS（格子数：約1.2兆点）におけるメモリ使用量180TB(当初：340TB)**

# 5. 開発状況 (2)通信コスト低減化

- ・ TDMA部分に**逐次演算とオーバラップ化による並列手法**を導入し、2段階転置を1段階に低減
- ・ **転置部分を近接ノード接続部分に割り当て**転置コストを最小化

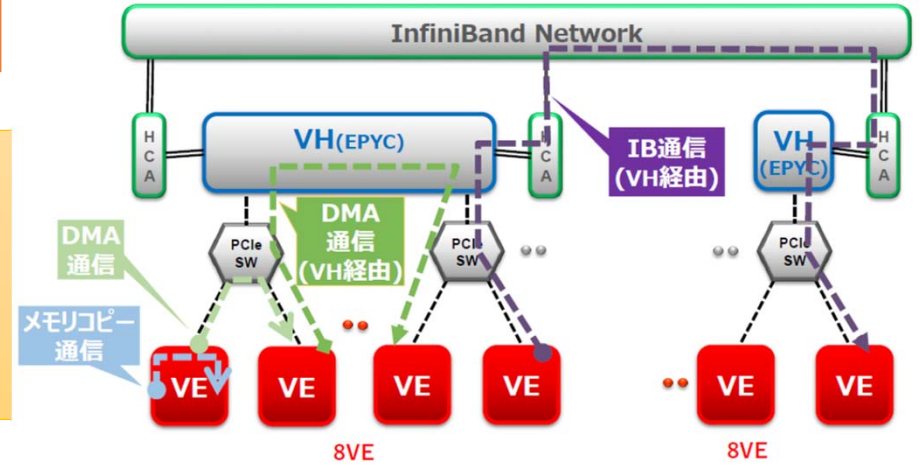
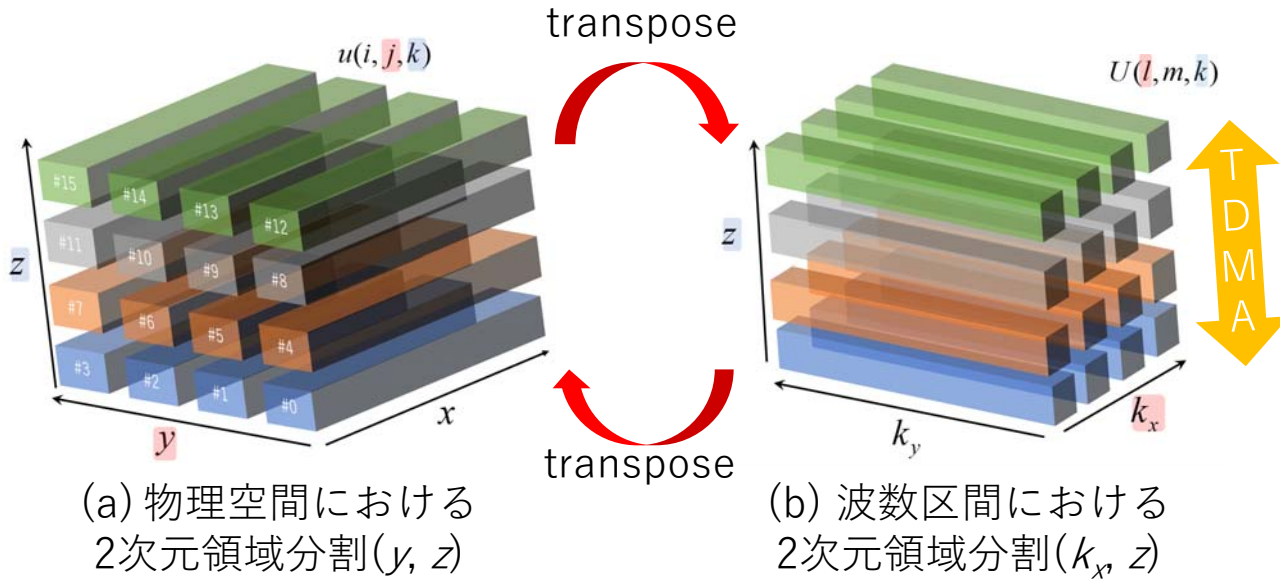


FIG. 東北大 AOBA-Aでのネットワーク構成  
 $y$ 方向領域分割(FFT)を**VH内8VE**に割り当て  
 $z$ 方向領域分割(TDMA)を**VH間**に割り当てる

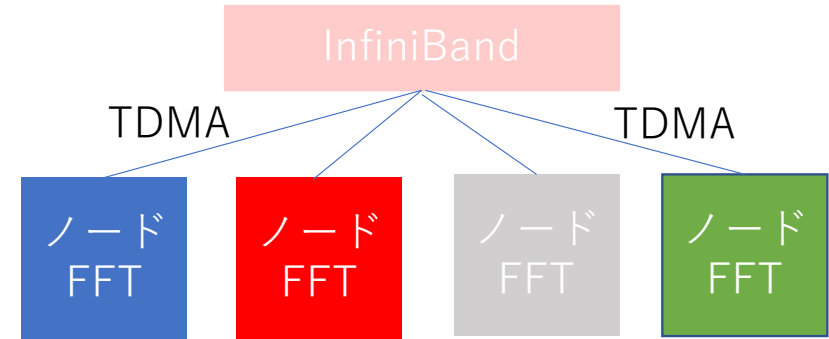


FIG. 京大Cray XC-40の場合  
 $y$ 方向領域分割(FFT)を**ノード内**に割り当て  
 $z$ 方向領域分割(TDMA)を**ノード間**に割り当てる



## 5. 開発状況 (3)演算効率の向上

(1)FFTライブラリの検討

- ・ FFTW, FFTE

(2)FFT演算部のOpenMP並列化

- ・ 多重1次元FFT or 1次元FFT

(3) FFTにおける演算と通信のオーバラップ化手法の開発



現在進行中

## 6. DNSコード開発の現況（まとめ）

### Re<sub>τ</sub>=16000規模のDNS（格子数：約1.2兆点）の必要要件と開発状況

#### [1] 低容量化

- ・要件：メモリ規模（200TB）で動作するための低容量化  
340TB→200TBへのメモリ量低減（**現状から約4割減**）  
→**180TB規模への低減化に成功**

#### [2] 超並列計算への対応

- ・要件：50万コア以上での並列計算が可能  
2次元領域分割コード → スペクトル法では**all-to-all通信コスト対策**が必要  
→**通信量を最小に抑えた2次元領域分割コード開発に成功**

#### [3] 高速演算

- ・要件：長時間時間積分を実行可能とするための、高速演算  
1200[Tflop/step] 3[s/step] → **400[Tflop/s] 級の実行演算速度**
- ・参考：チャンネル流における最速コードは**180[Tflop/s]** Lee et al. SC2013  
→**演算効率の向上を実施中**  
(実際の評価は、次期東北大システムが必要、本年度は場合によっては他システムでの評価を実施予定)