



SCALE Project

気象気候惑星科学共通ライブラリーを目指して

富田 浩文

(独) 理化学研究所・計算科学研究機構
複合系気候科学研究チーム

& Team SCALE

(Scalable Computing for Advanced Library and Environment)

お話しの内容

本日は、NICAMのお話はしません。

- 背景:
 - モデルづくりの面から、
 - 計算機トレンドの面から、
- What is Team SCALE?
- SCALE project 目指しているところ
- 現在取組中のこと
 - スカラーマシン最速LES
 - マイルストーンとなるサイエンスターゲット

背景

- おのおののグループで、おのおのモデルを作っている。
 - モデルを作ること自体はとてもいいこと。
 - モデルを作ることによって、いろいろ原理的なことが分かる。
 - でも、力は、分散。
 - 分散するのもいいこと？
- そもそも、地球科学のモデルは、一人で作れるものではない。
 - コンポーネントが多く、また多様化している。
- 多くのコンポーネントの組み合わせによって、結果が異なる。
 - 追試必要だが、やってるひまない。
 - 不健全。
- コンピュータのアーキテクチャが複雑化しようとしている。
 - ヘテロジニアス、多階層メモリ構造
 - 普通に書いていたのでは、性能がでない。

一つのコンポーネントでも腐るほどある

- 流体力学部分：
 - 水平離散化：
 - スペクトル法、ダブルフーリエ、
 - 正20面体(各種)格子、
 - 立方体格子、陰陽格子、スペクトルエレメントなど、
 - 鉛直スキーム
 - 静力学方程式系
 - ほぼ、プリミティブ方程式できまり。
 - 各種非静力方程式系
 - 非弾性系、弾性系
 - 山の取り扱いなど
 - 時間スキーム
 - Explicit
 - Time split : Leapfrog, RK2,3,4
 - Implicit
 - 鉛直方向のみ? 3Dで?

物理過程の複雑化、多様化

- 各種パラメタリゼーション
 - 対流スキーム
 - Kuo, AS, Tiedike, K-F, chikira, yoshimura
- 各種雲微物理
 - 大規模凝結スキームも多々
 - 1モーメント→ 2モーメント→ ビン法
- 放射：
 - 平衡平板モデル→ 3D放射
- エアロゾル、化学パッケージ
 - 種類がどんどん増えている。
- 陸面過程、などなど
 - 評価方法よくわからない。(私の勉強不足)
- 海洋モデルの渦パラメタリゼーション
 - GM他
- 生態系、氷床、、、う～ん、

計算機構・気候チームからの提案

- 多くの主要なスキームを網羅し、
 - **誰でも自由**に使えて、
 - **性能が出せる**ライブラリーを提供したい。
 - モデラーの育成
 - エクサスケールそれ以上を意識。
 - オープンソース
 - **ドキュメント整備**も大事。
- 一つのグループでは無理。
 - 国内外多くの機関、グループを巻き込みたい。
 - MIROC + NICAM + 気象研・気象庁(?) + 神戸大・京大
 電脳グループ + JAMSTEC・ESCグループ + 名古屋大
 CReSSグループ

- ライブラリー構築はいいが、、
- HPCの動向を考えとかないと、、
- 使えない！

さて、HPCの動向

- スカラー超並列
 - 地球シミュレータ1 640 -> K computer 80,000
 - エクサ機は? 10,000, 100,000, 1000,000
 - 壊れる!壊れる!壊れる!
- メニーコア化(コモディティベース)
 - AMD opteron, Intel sandy (Ivy) bridge
 - Knight, MICなど
- ヘテロジニアス化
 - GPGPUなどアクセラレータ装備(例:TSUBAME2)
- 主記憶メモリー性能の相対的低下
 - ES1: 4 -> ES2: 2.5 -> K: 0.5 -> エクサ:0.1???

避けられない二つの方向:
ベクターからスカラーへ
汎用機から専用機へ
電力制約から

● 深刻なBF比の低下:

- ✧ Peak: 10PFlops, B/F: 0.5 (K Computer system)
 - ✧ performance efficiency
: 6-12% for double precision = ~0.6-1PFlops
- ✧ Peak: 20PFlops, B/F: 0.2
 - ✧ 3-6% = ~0.5-1PFlops
- ✧ Peak: 100P~1EFlops, B/F: 0.1?
 - ✧ 2-3%...

うちらにとって好ましくないこと

- 複雑化するアーキテクチャ
 - ヘテロジニアスアーキテクチャ:
 - アクセラレータどう使う？
 - うまく使えれば、とてもうれしい。
 - ASUCAなど
 - 主記憶バンド幅低下:
 - 流体計算は全滅なのか？
- BFあげることを要求するの？
 - 電力問題
- 大量IO処理:
 - ポスト処理も並列に
- フォルトトレランス
 - いつもどっか壊れている。
 - アプリ側でチェックポイントを強いられる。

これらの根本的対処を考えるのがAICSの業界での立ち位置か？

計算機構の役割

- いろんなグループのモデルコンポーネントの集約・ライブラリー化。 [SCALE project](#)
 - 力点：
 - ポストペタでのスケール
 - 対障害性問題
 - 新しい並列言語体系
 - ポストMPI
 - 基本、オープンソース
- 各モデルグループからのコントリビューションを整備、オープンソース改変可能型ライブラリーの提供
- AICS内、気候チームだけではなく、他計算科学チーム、計算機科学チームとの密な連携のもとに行う。

- Transdisciplinary Research Team in RIKEN AICS



**Computational
Climate Science
Research Team**

TL : H. Tomita



**System Software
Research Team**

TL : Y. Ishikawa



**Programming
Environment
Research Team**

TL : M. Sato



**Co-design of application, library, and environment
for weather/climate research**

まず、やること(本格LIB構築の前に)

- 将来の気象・気候モデル(典型的)を想定して、いったいKコンピュータでどのぐらいの計算ができるのか？
- 次世代へ向けた可能性調査研究が必要
 - 問題はどこか？
 - アルゴリズム改変は必要か？
- SCALE-LES project

SCALE-LES の思想

✻ まずは、とことん！

- ✻ Kコンピュータ上でパフォーマンスを最大限引き出す。
- ✻ **より簡単に, より速く, 精度は必要かつ最低限**
 - ✻ FLOPSを負うのはもはやナンセンス。
 - ✻ やりようによっては、いくらでも計算インテンシブにできてしまう。
 - ✻ E.g. ビン雲微物理最大限精緻化して、毎ステップ計算させる (×)

✻ 気象学での発展的意味

- ✻ 非常に高領域で行う。
= 最初の “mesoscale LES” model

✻ 計算方法はなにがいいのか？

- ✻ Full-explicit scheme : 通信は隣接のみ
- ✻ HEVI, HIVEなどラインナップをそろえ、正当比較
- ✻ Final goal : 知見を活かし、ライブラリー構築へ。一部、コンパイラやシステム屋に提言。

Overview of SCALE-LES

Main Model components	
Dynamics	Full-explicit, non-hydrostatic system (Miura and Miyamoto)
Cloud Microphysics	Double-moment 6 water category (Seiki)
Sub-grid scale	Smagorinsky-type scheme
Radiation	MstrnX (Sekiguchi & Nakajima)

- 等方格子でやる！ → directional splitはしない。する意味がない。
- ロードストアの軽減を念頭に→ 時間スプリットもしない

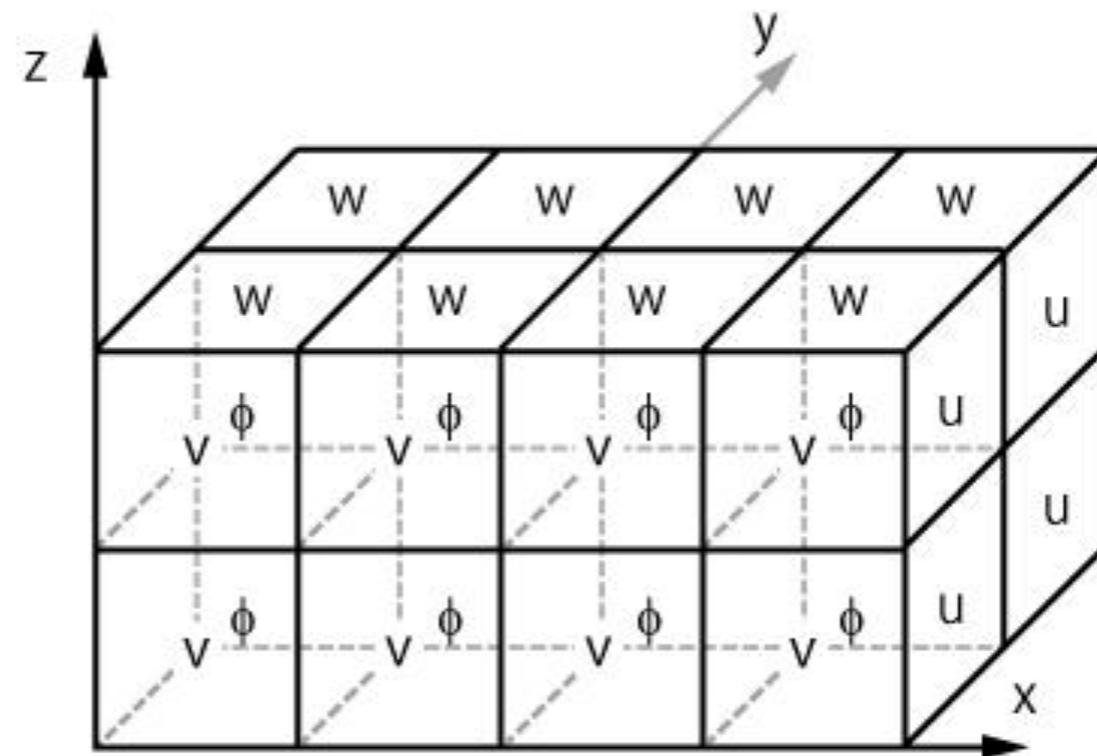
手法

● 力学過程:

- 3DFull Explicit NS方程式を解く
- 格子系:カーテシアン格子
- 時間積分スキーム:3次Runge-Kutta
- 空間差分:4次中央差分
- その他トレーサー(水蒸気など)
 - 非負保証のため、FCTスキーム

● 物理過程

- LES乱流モデル:大気成層を考慮したSmagorinskyモデル
- 雲微物理:ダブルモーメントスキーム(Seiki 2011)、ビン法(Suzukiモデル)
 - 雲微物理過程の理論式に基づいた定式化
 - 雲粒子の粒径の正確な予報
 - 低コストで、
 - チューニングパラメーターの削減
 - 雲光学特性の高精度化
- 放射: MSTRNX(Sekiguchi & Nakajima 2008)
 - NICAMより導入



まずは、いろいろテストケース

- 力学テストケース

- 移流テスト: CWCをみたすように、非負を保証するように
- 乱流テスト:
 - とりあえず、追試

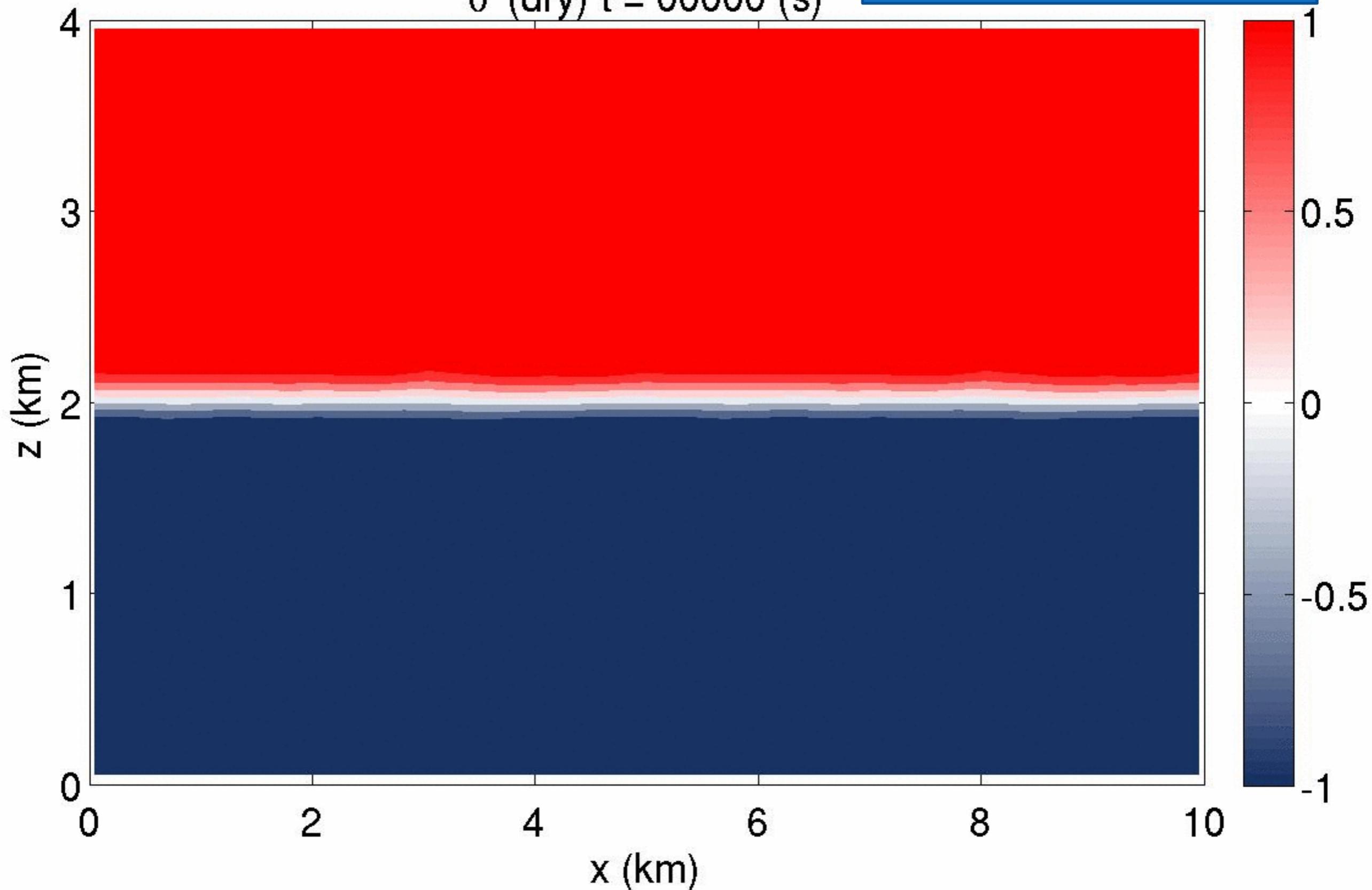
- 雲関係:

- 1モーメント、けすらー
- 2モーメント Seiki scheme
- Bin法
- テストケースは、
 - Warm bubble, super cellなど

KH不安定テスト

宮本佳明 (AICS)

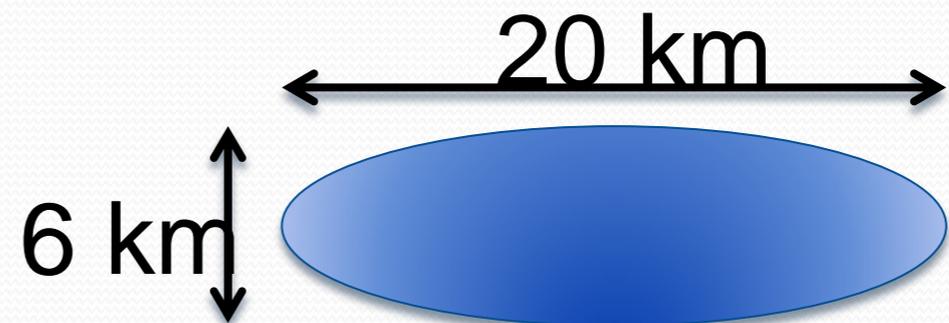
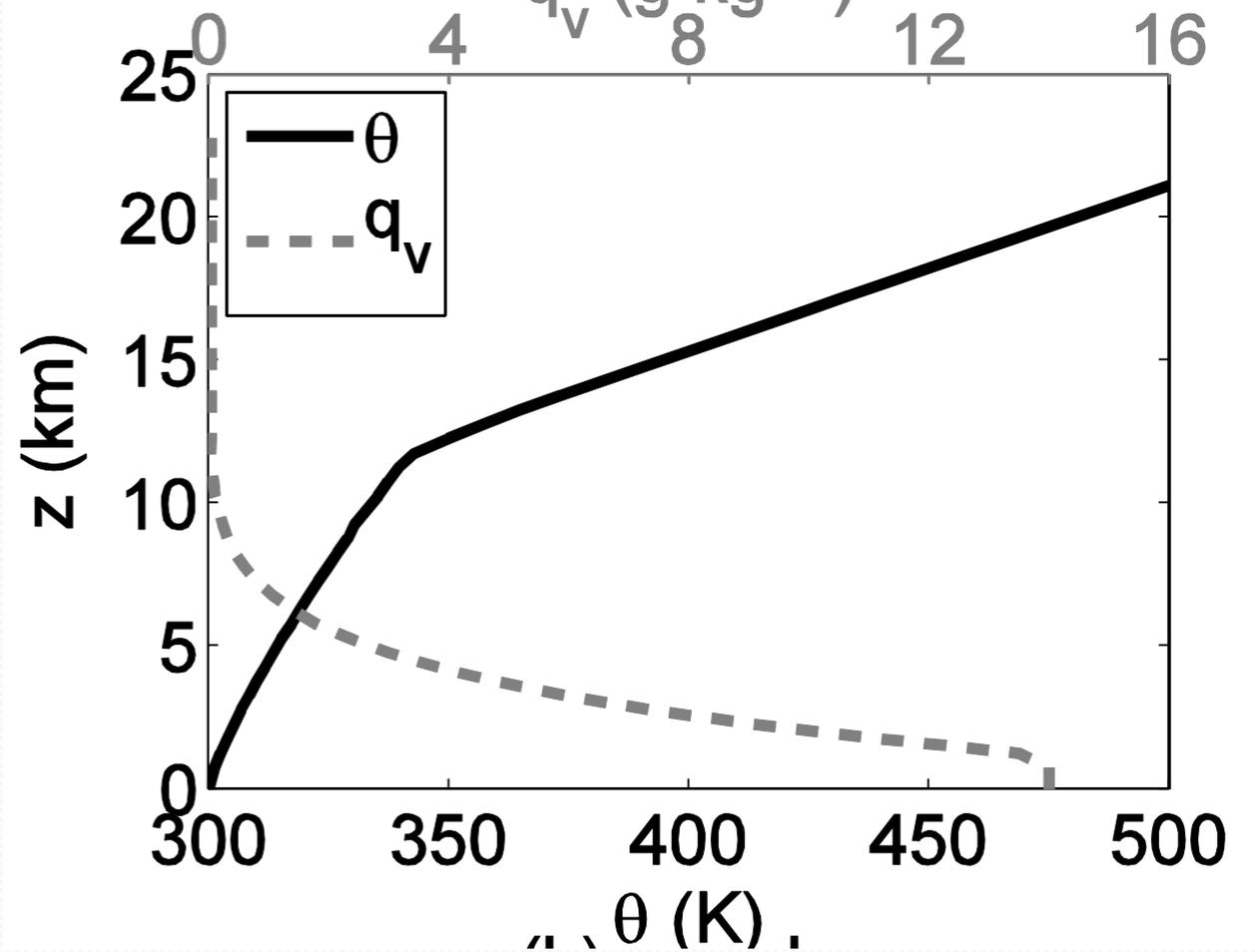
θ' (dry) $t = 00000$ (s)

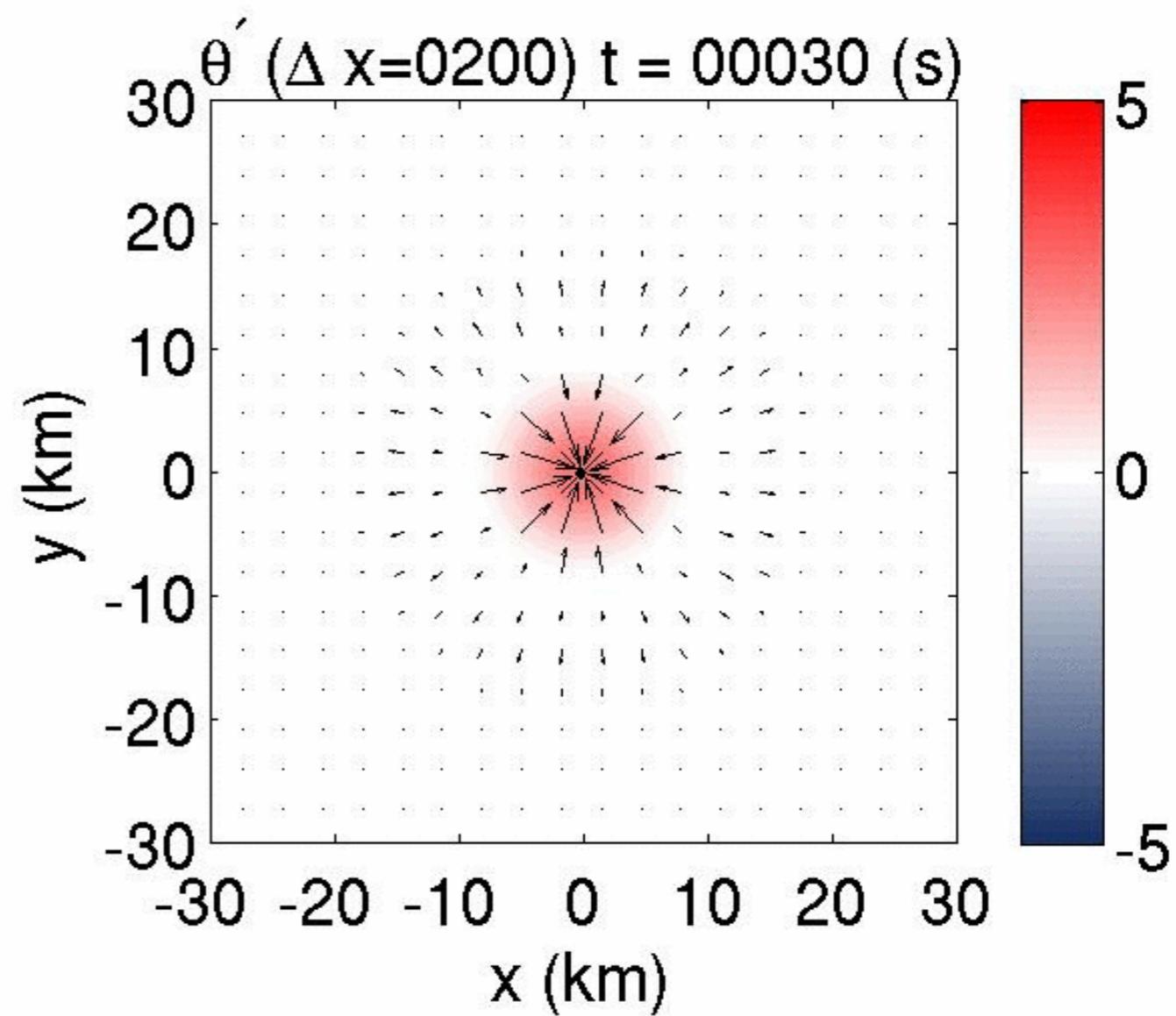
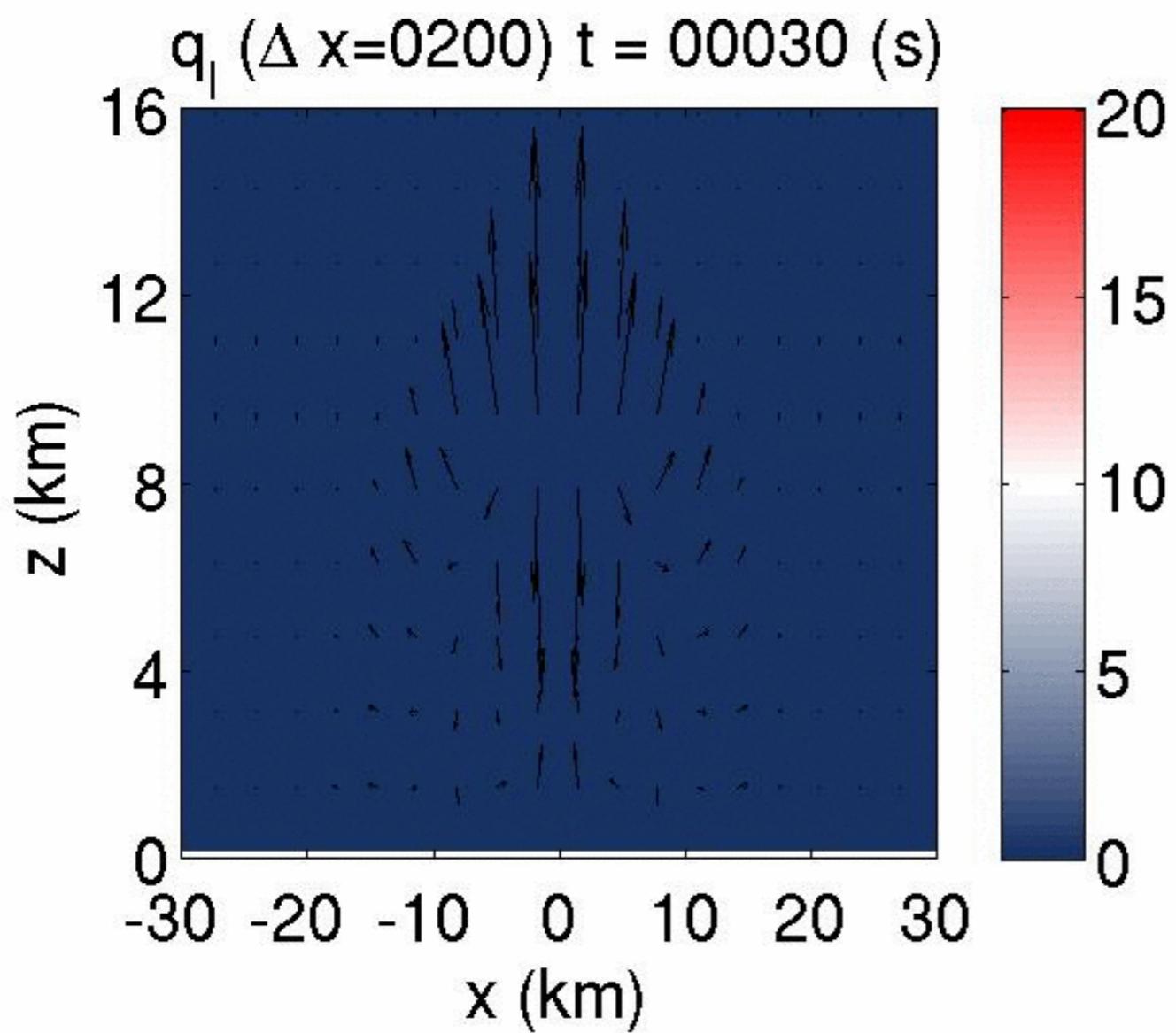


Warm bubble実験

宮本佳明 (AICS)

- Ooyama (2001)
- 初期環境場 (図1、温位・水蒸気): Jordan (1958)
- 長径10 km、短径3 kmの楕円温位偏差 (最大値 5 K)
=> 領域中心・高度 0.5 kmに置く
- 計算領域: 80 x 80 x 20 km (強安定層 > 15 km)
- 雲物理: Kessler (1969) warm rain (氷相無し)

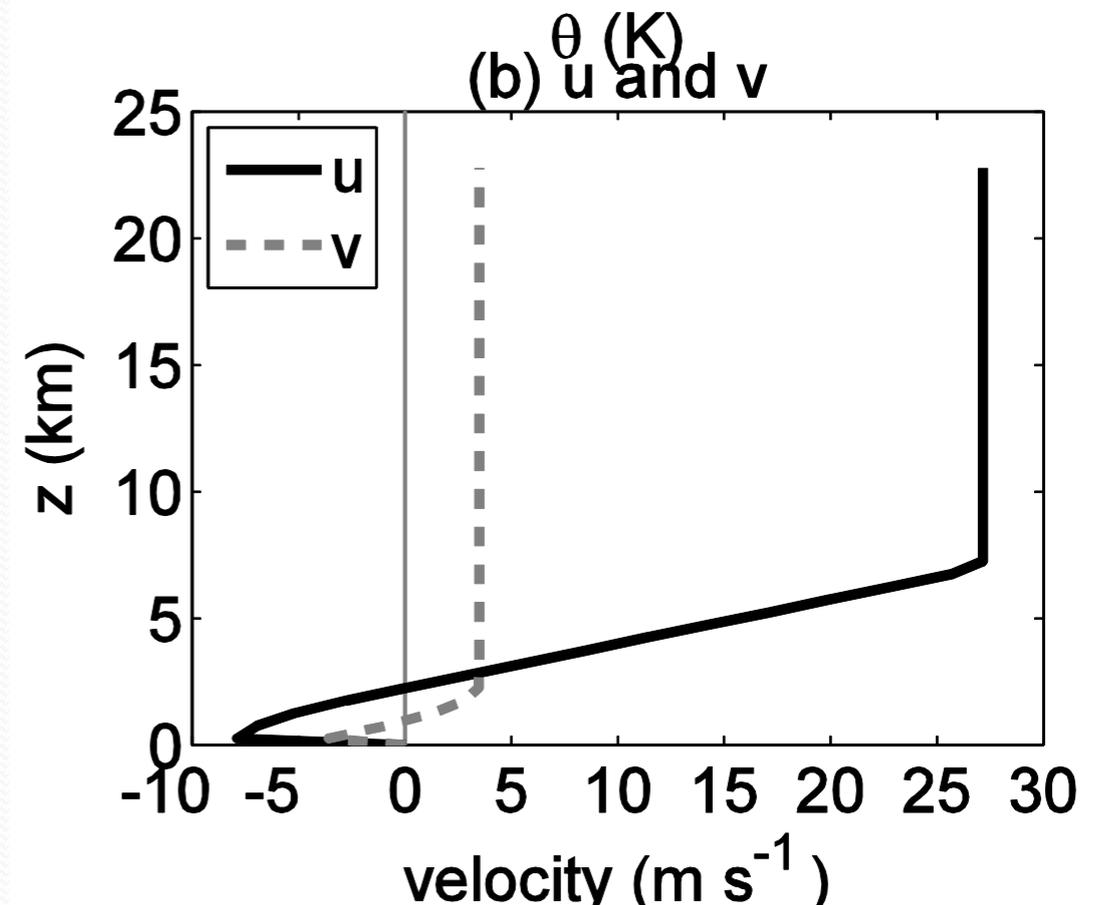
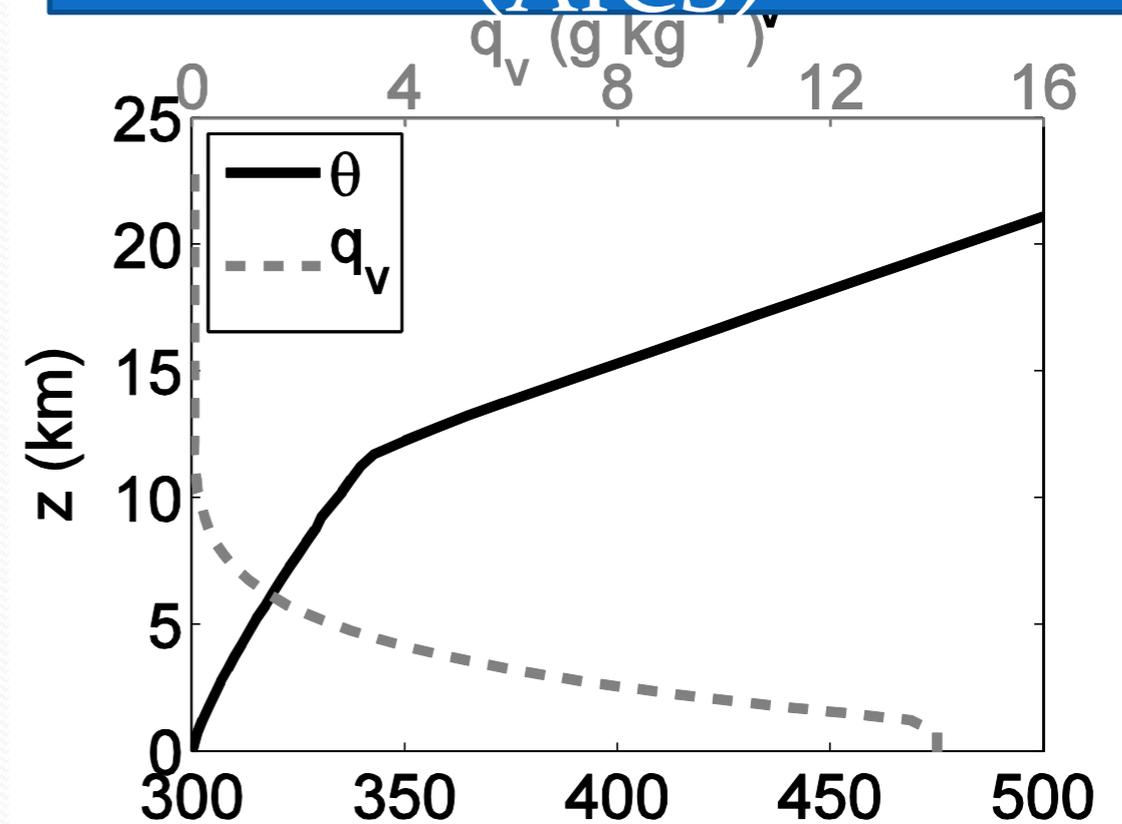




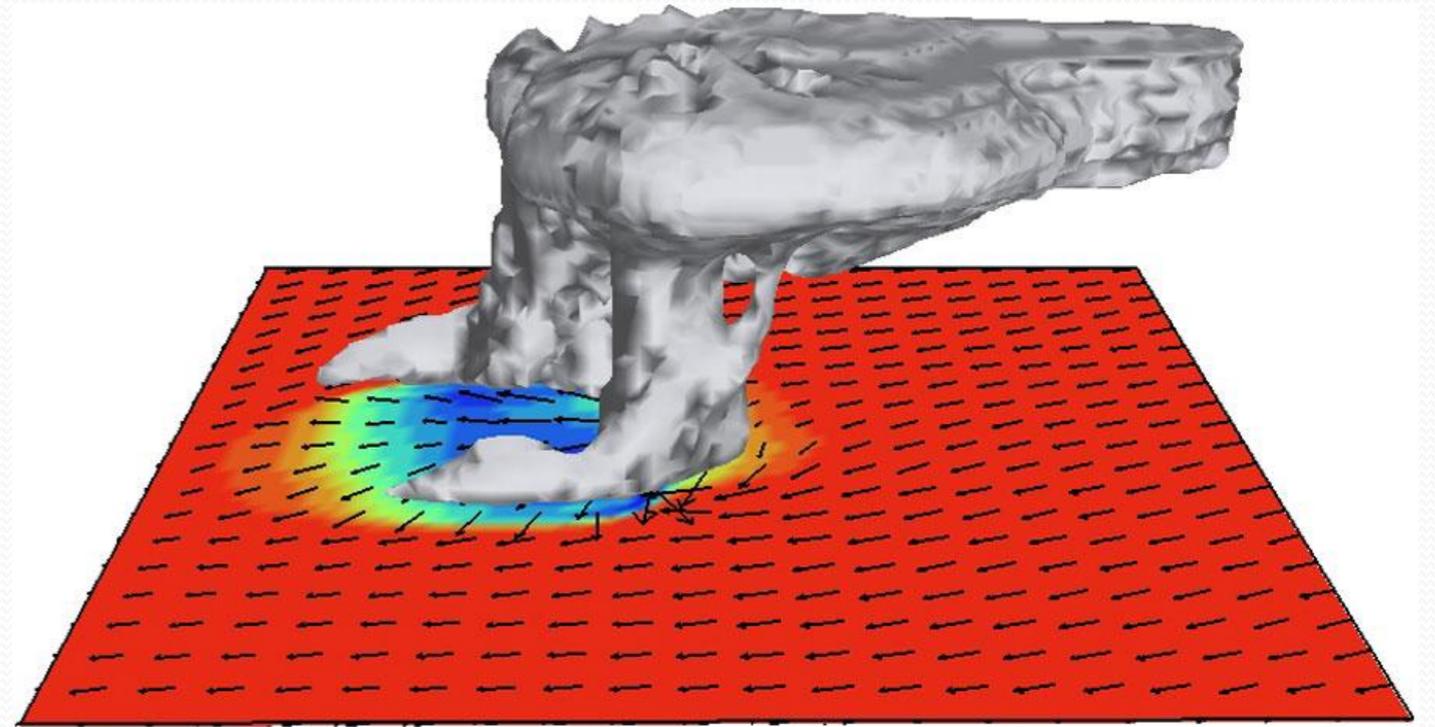
Super cell実験

宮本佳明・吉田龍二
(AICS)

- Counter part のWRFモデル実験と同一の実験
- 長径10 km、短径3 kmの楕円温位偏差(最大値 5 K)を領域中心・高度 0.5 kmに置く
- 水平風の鉛直シアー(図2)
- 計算領域: 160 x 160 x 20 km (強安定層 > 15 km)
- 雲物理: Kessler (1969) warm rain (氷相無し)



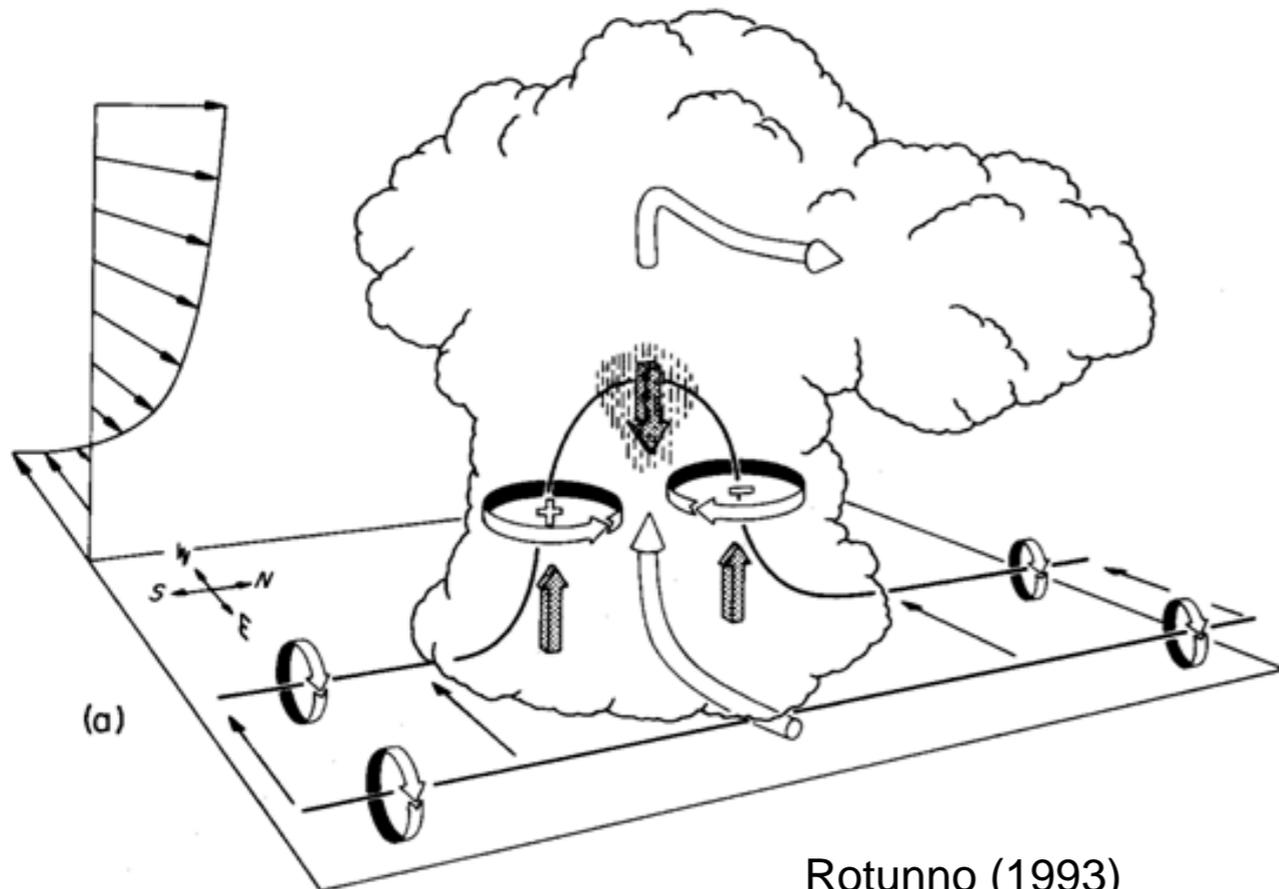
supercell



<http://www.mmm.ucar.edu/asr2000/supercell.jp>

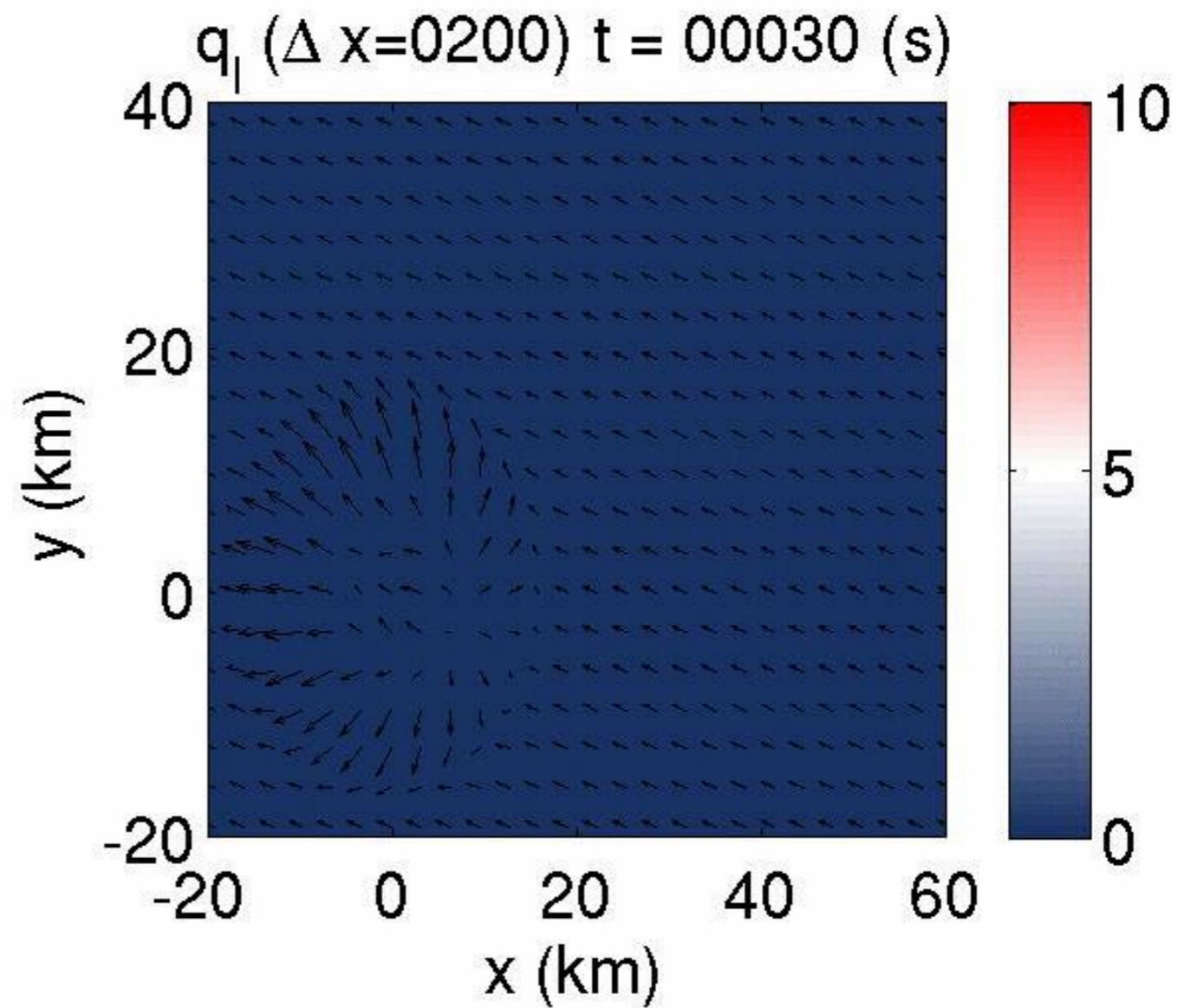
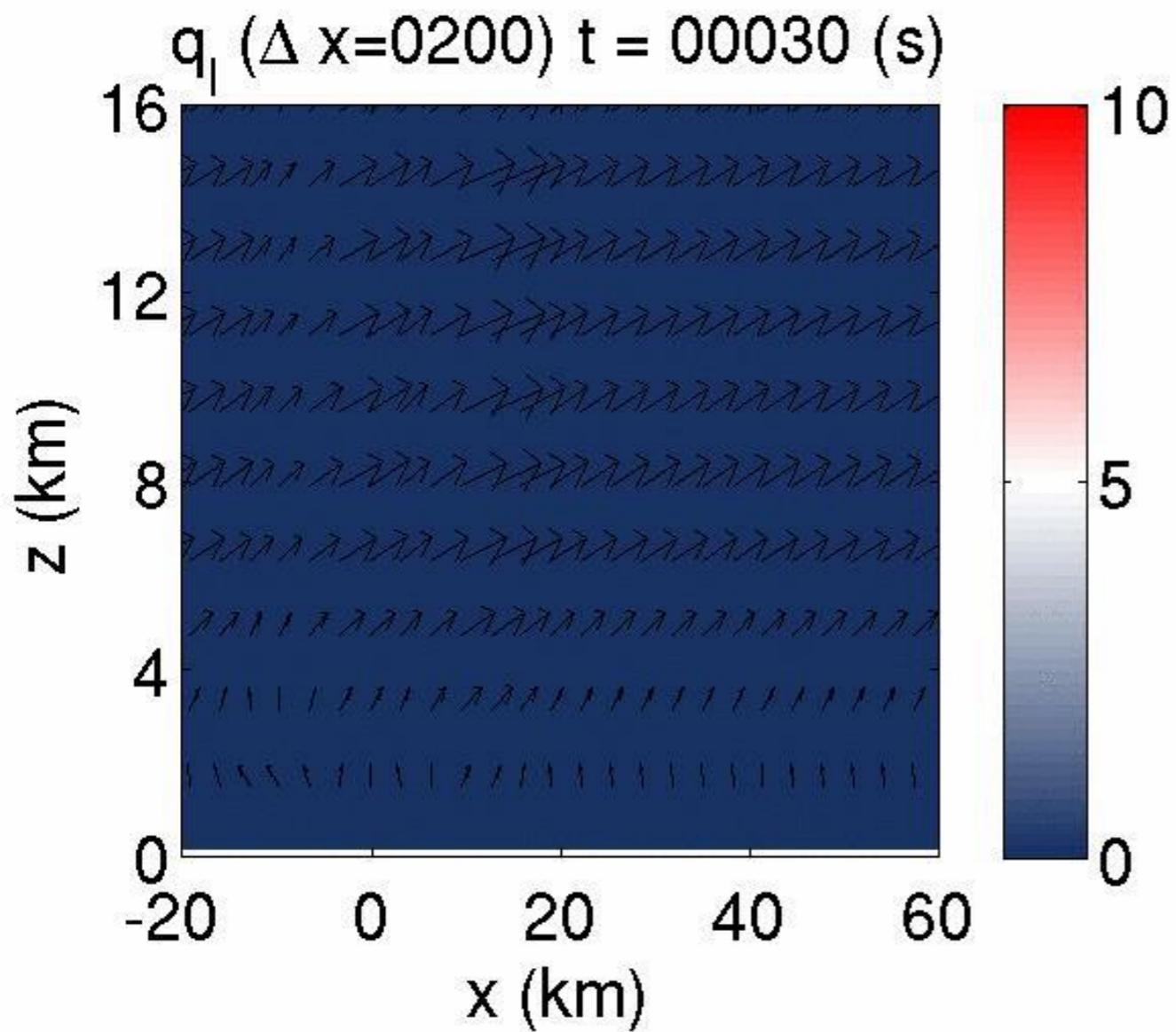
9

Scamarock 提案のテスト

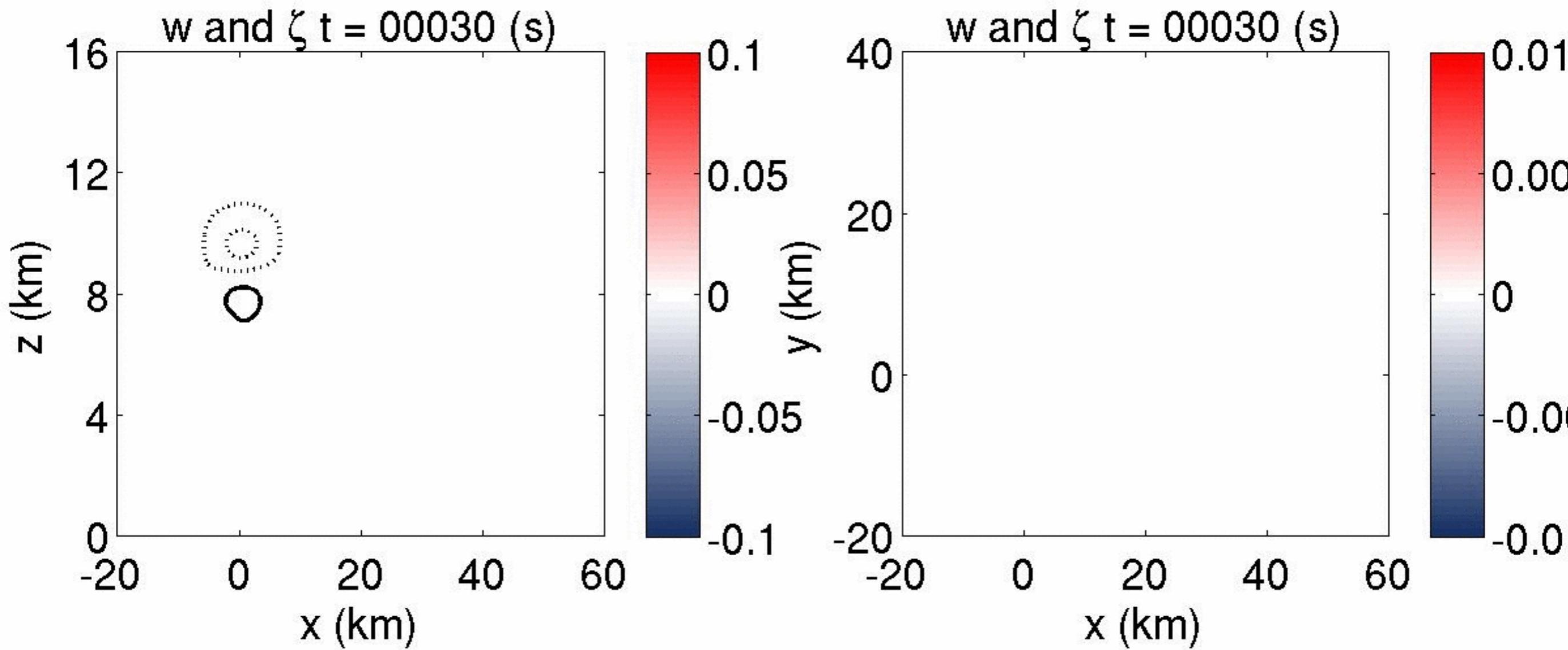


Rotunno (1993)

10. Super cell実験



10. Super cell実験

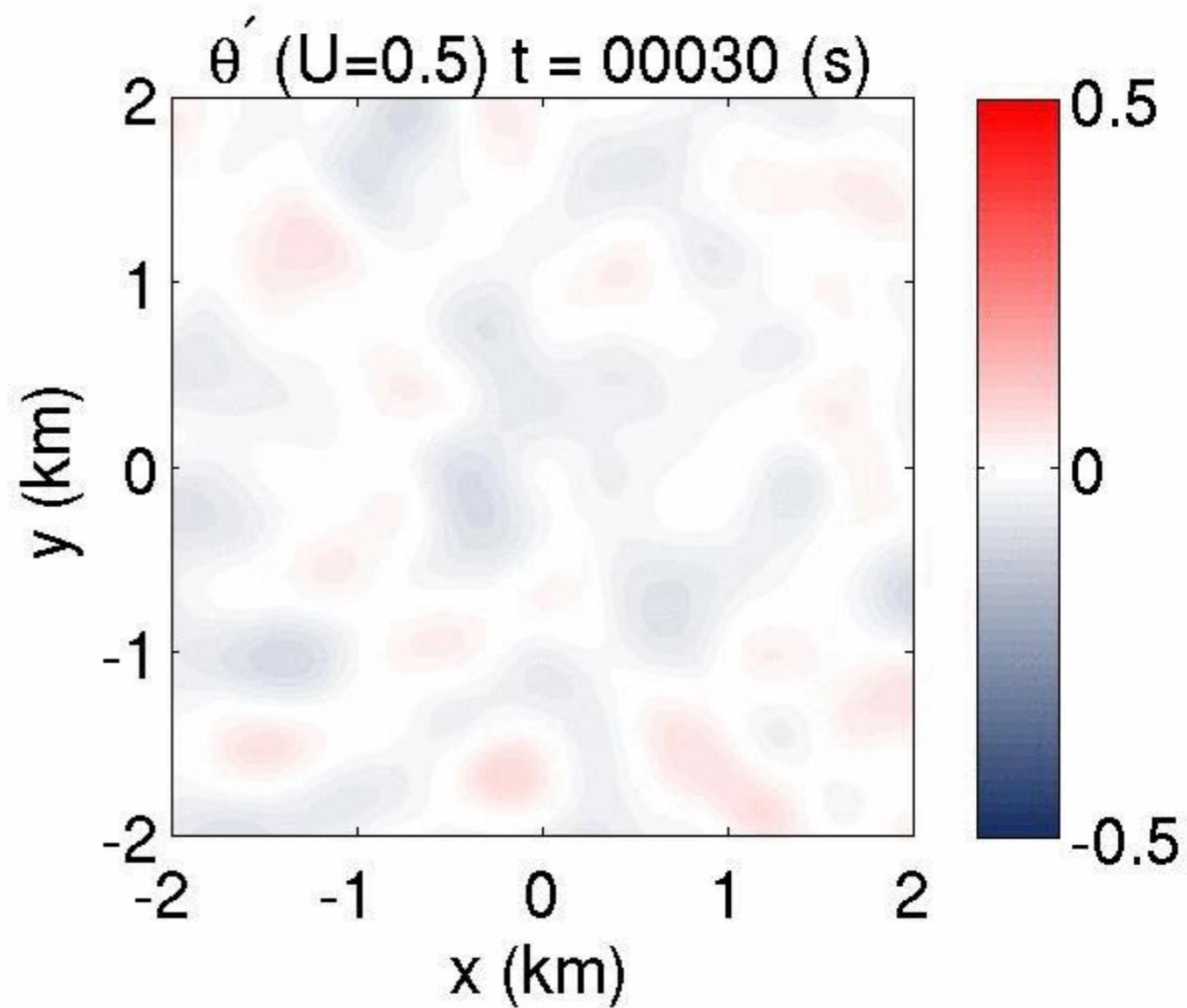
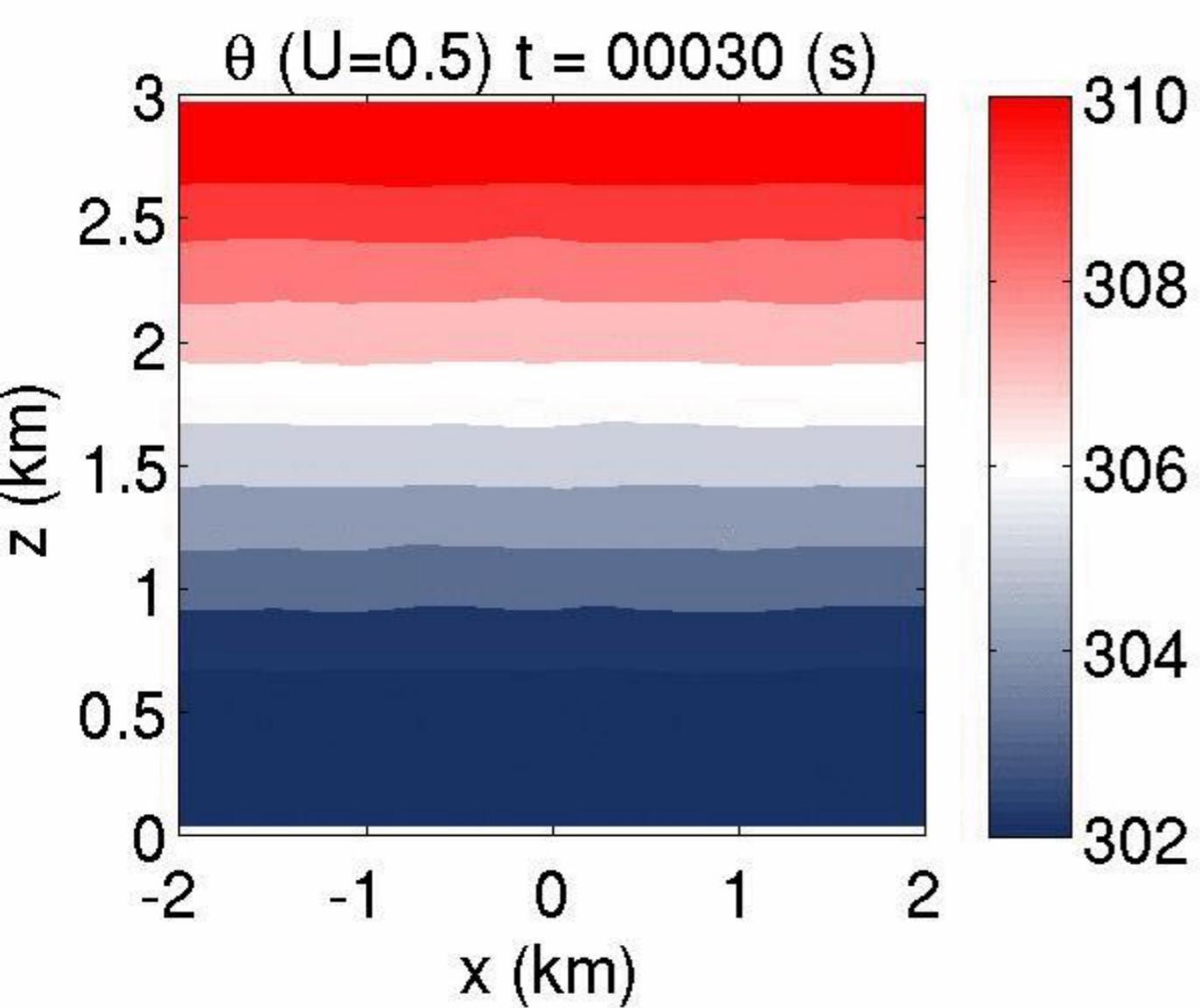


乱流実験

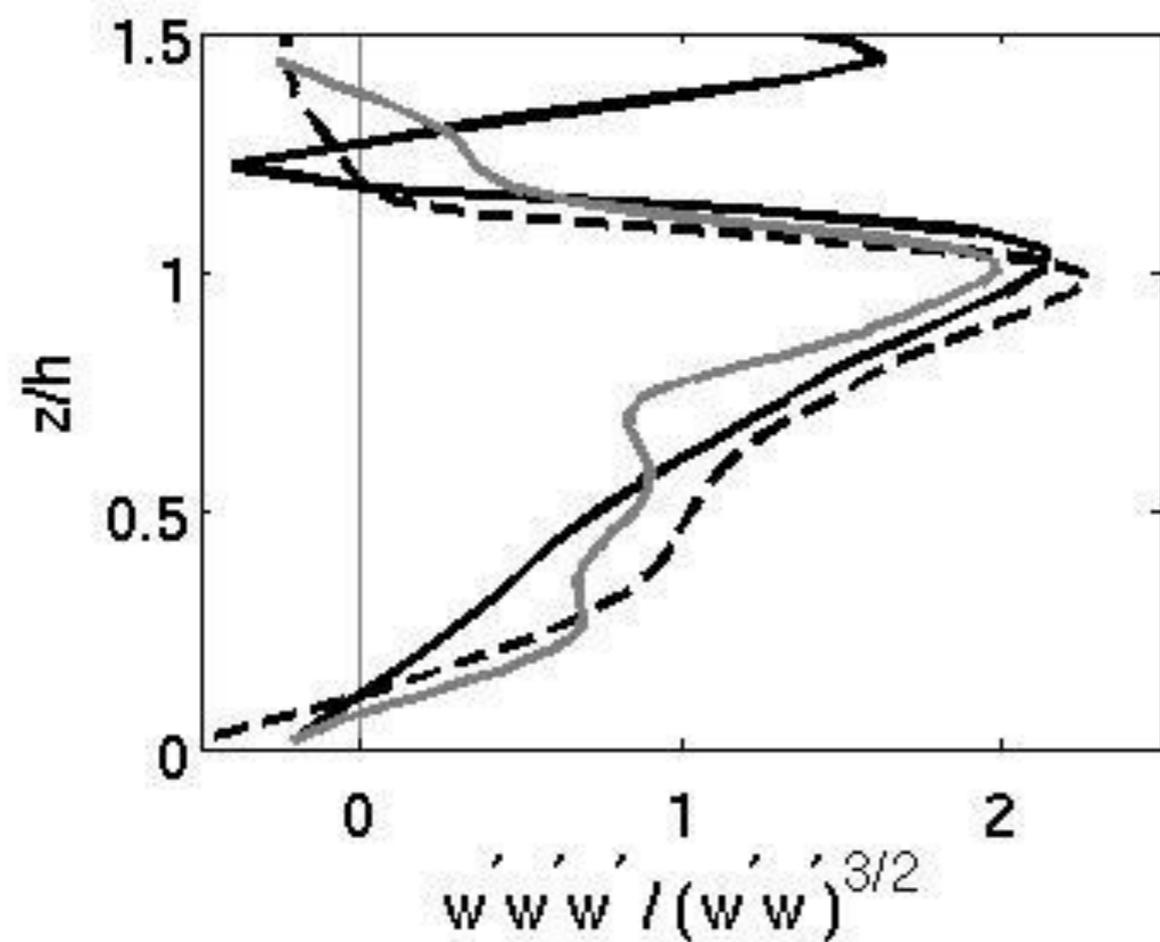
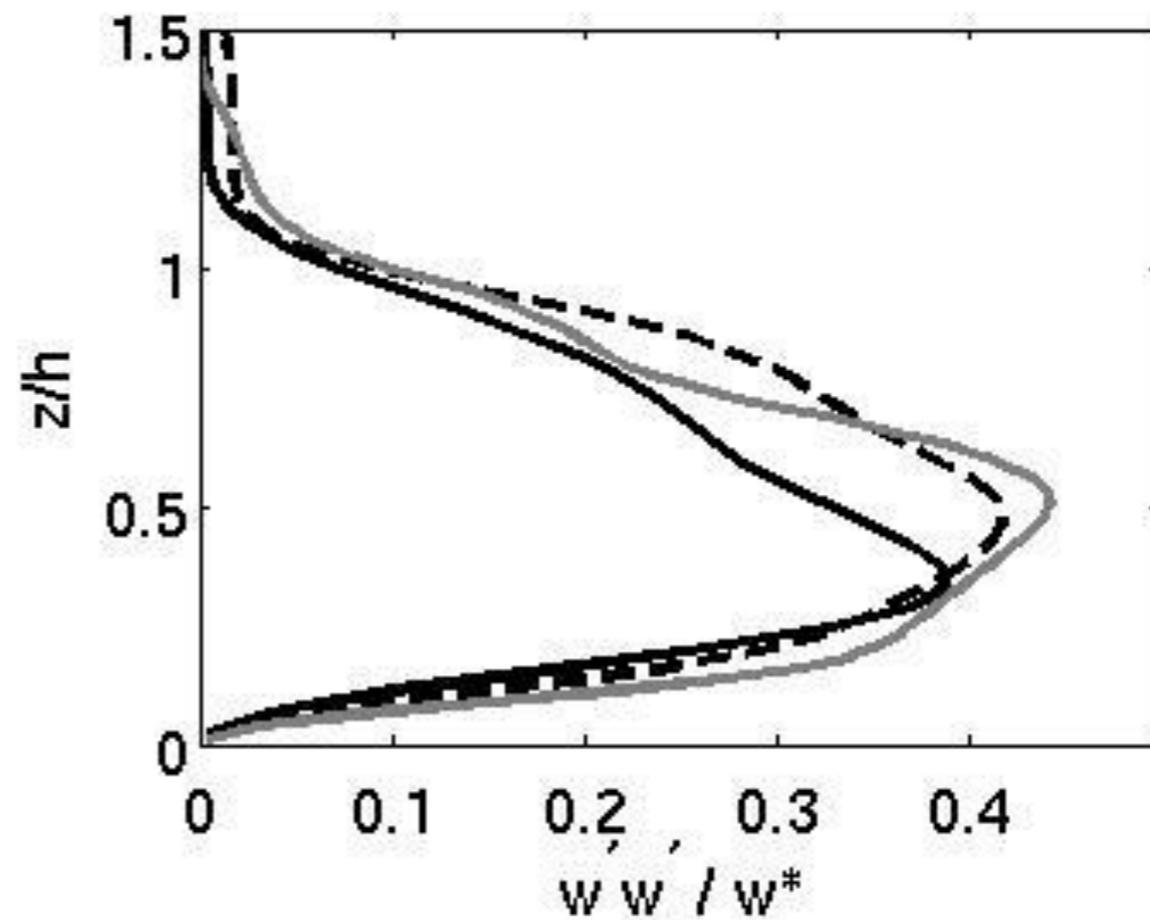
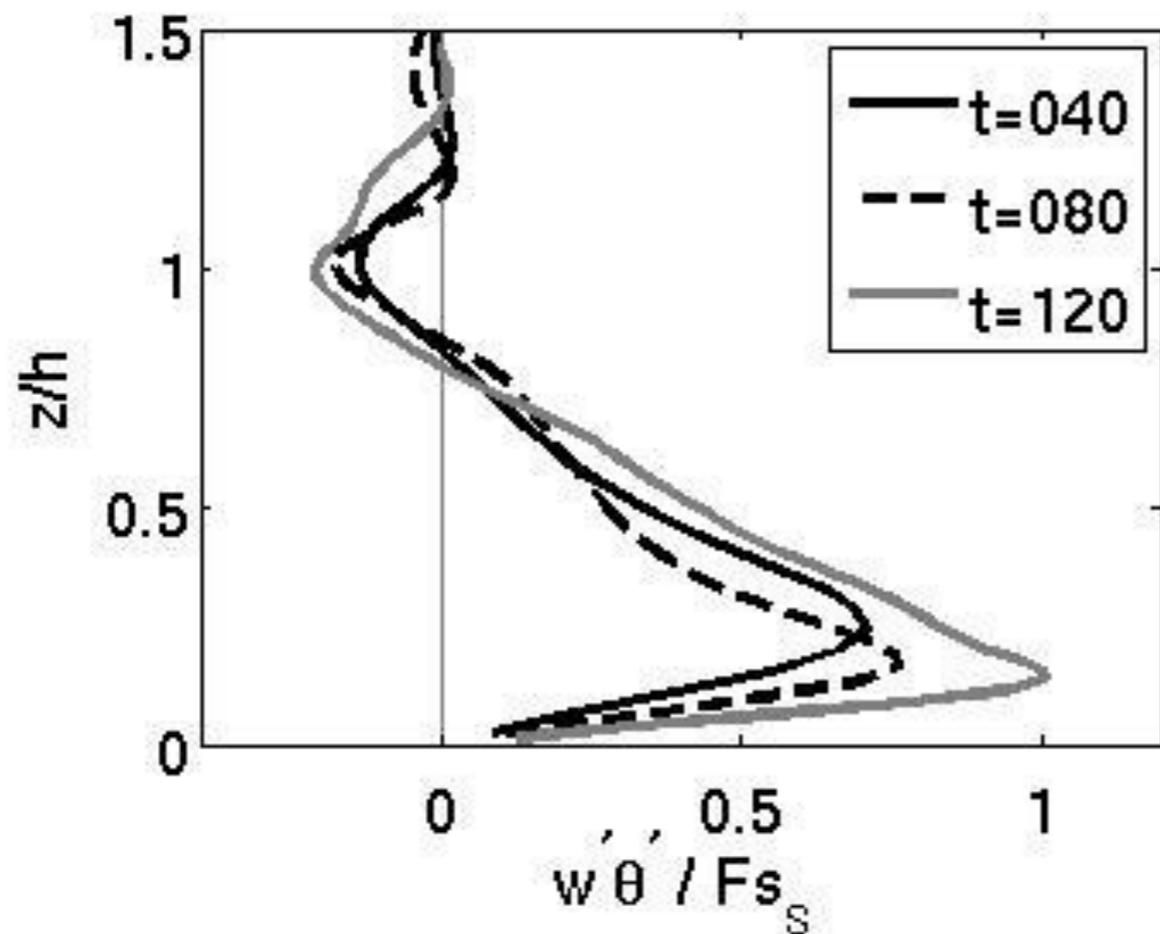
宮本佳明(AICS)

- 新野研の検証実験と同じ設定(気象研究ノート参照)
- 計算領域: $4.5 \times 4.5 \times 3$ km (弱安定層 > 0 km)
- 温位擾乱(振幅1 K)
- 水平風:鉛直シアー無し
 - $U = 0.5 \text{ m s}^{-1}$
 - $U = 5.0 \text{ m s}^{-1}$
 - $U = 15.0 \text{ m s}^{-1}$
- 境界から、下面熱フラックス与え続ける(0.2 K m s^{-2})

Dry, $z = 25$ m



12. 乱流実験

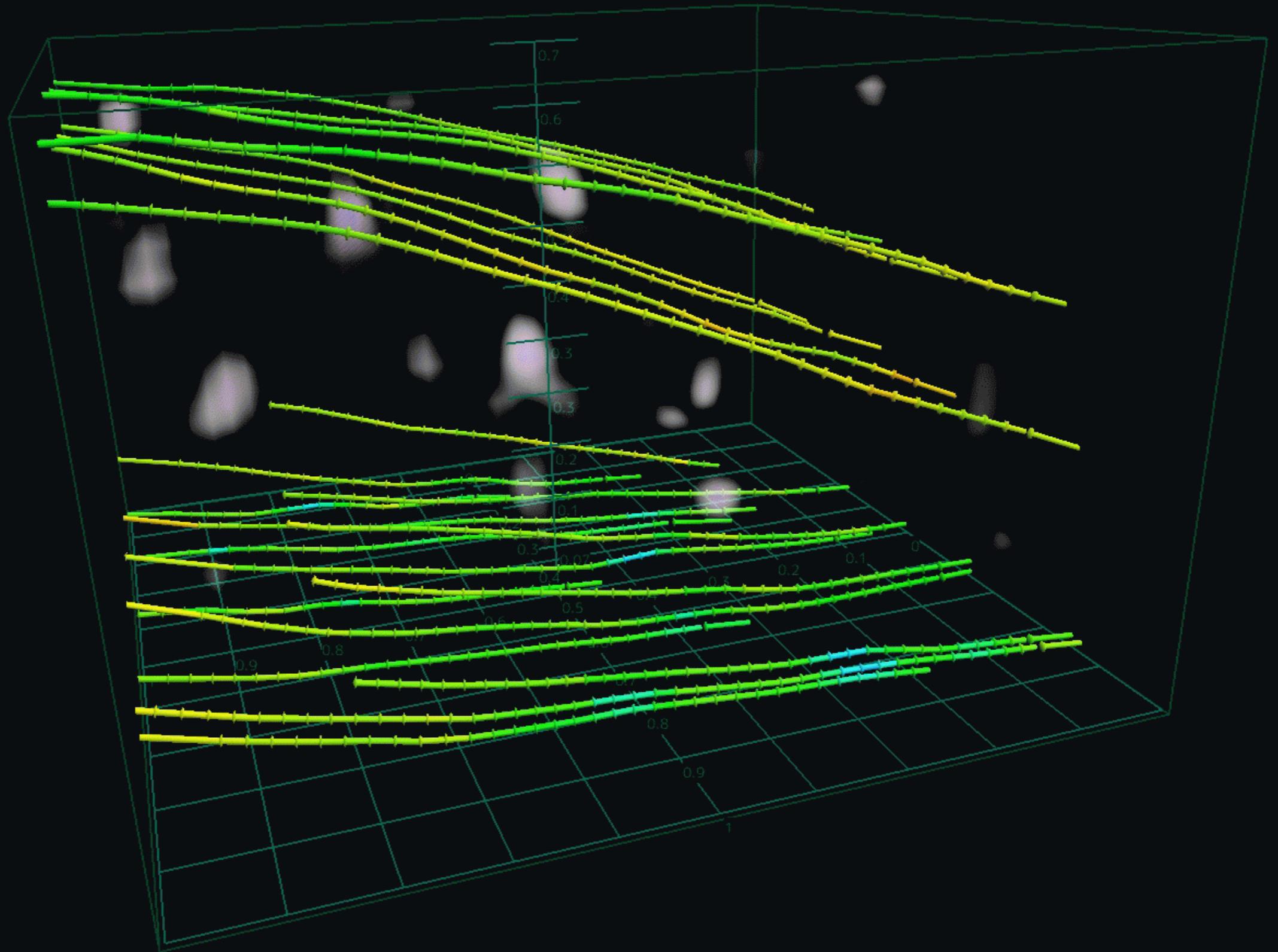


※全ての変数は水平平均値
($'$ は水平平均からの偏差)

h : 境界層深さ
(=浮力フラックスの極小高度)

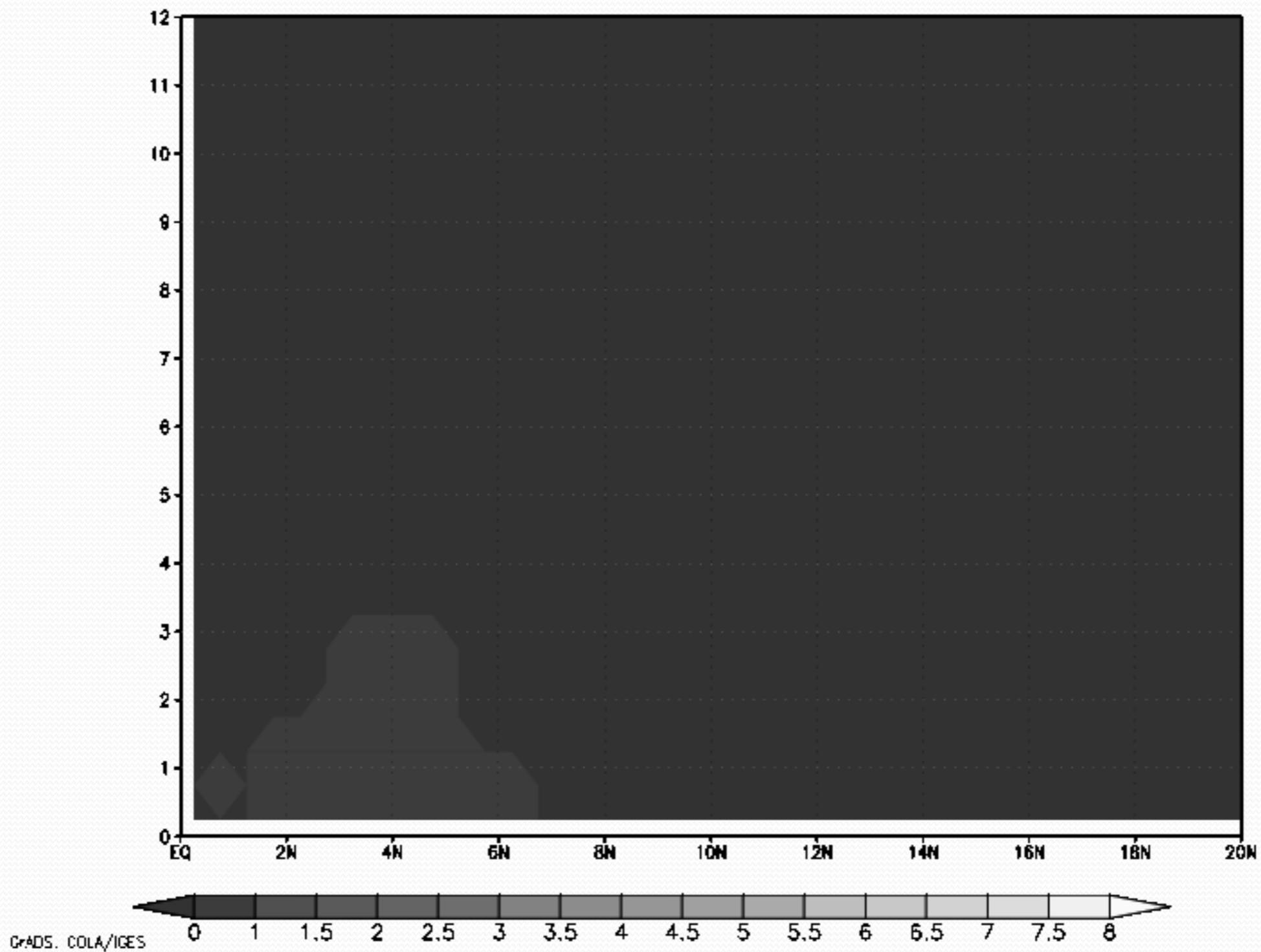
w^* : 鉛直速度スケール

F_s : 底面顕熱フラックス



ベンはじめました。

佐藤陽助 (AICS)



で、どこかで全系計算をやりたい。

- サイエンスターゲット：
 - 層積雲の崩壊過程のメカニズム
解明

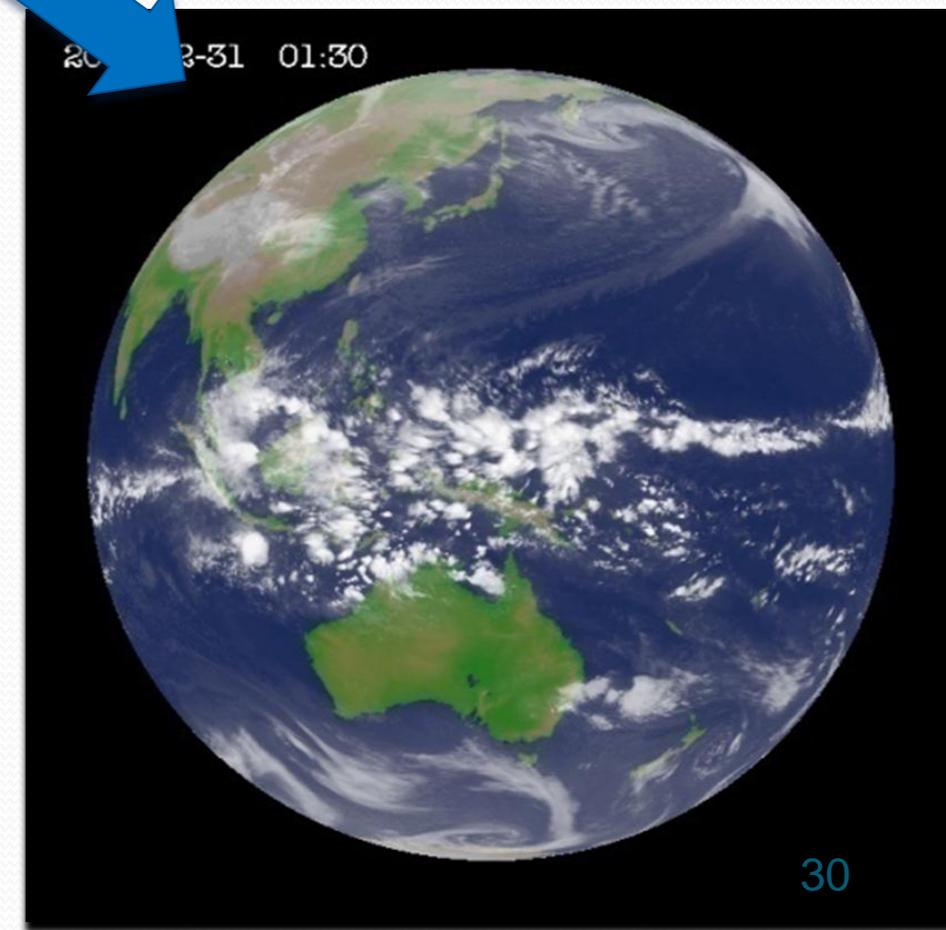
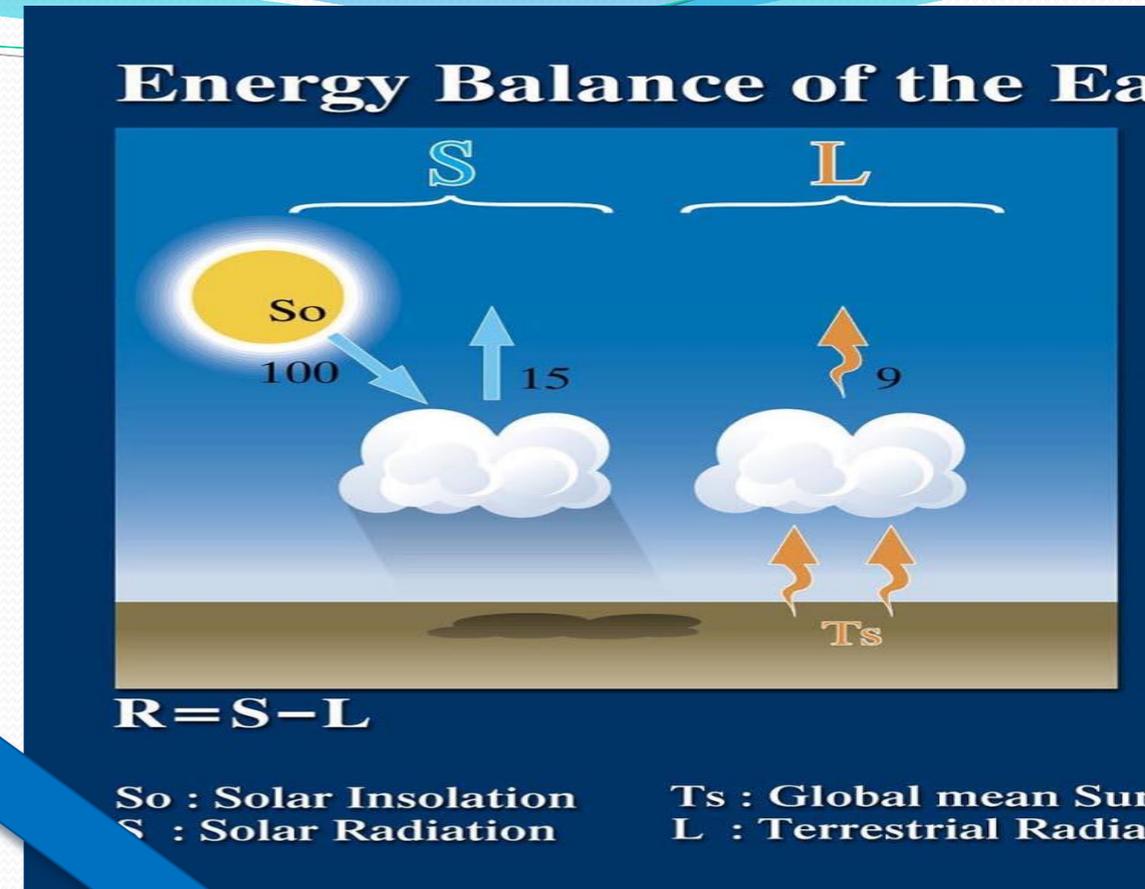
気候における雲の重要性

気候における雲の役割

- **日傘効果と温室効果**
- **背の高い雲(積乱雲) = 温室効果卓越**
 - 熱帯域での積乱雲は大気のエンジン
 - NICAM(全球雲システム解像モデル)により、直接解像できている。

• 背の低い雲(層積雲、積雲)の重要性 = 日傘効果卓越

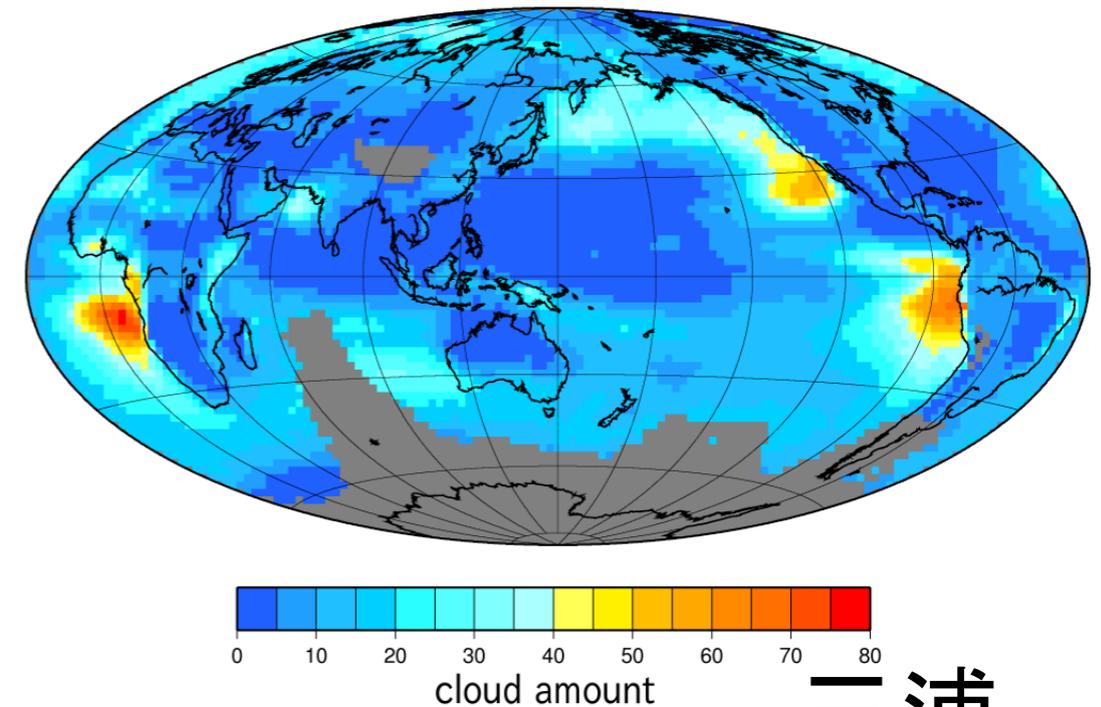
- 鉛直、水平スケールが小さい！
 - **まだまだ、全球モデルでは直接解像無理！**
- **パラメタリゼーションに頼らざるをえないのが現状**
- **そもそも、生成消滅の理解が乏しい**



背の浅い、小さな雲群の気候に与えるインパクトの重要性

- 層積雲
 - 亜熱帯・東海岸で発現
被覆率: 50 %以上
- 気候に与えるインパクト大
 - 特に日傘効果大
- 時間・空間スケール大だが、
 - 現象を記述する
物理過程は非常に細かい
- 熱帯へかけて層積雲の崩壊過程
 - 層積雲から積雲への遷移の
メカニズムが分かっていない!

Cloud Amount (1984–2007 Aug)



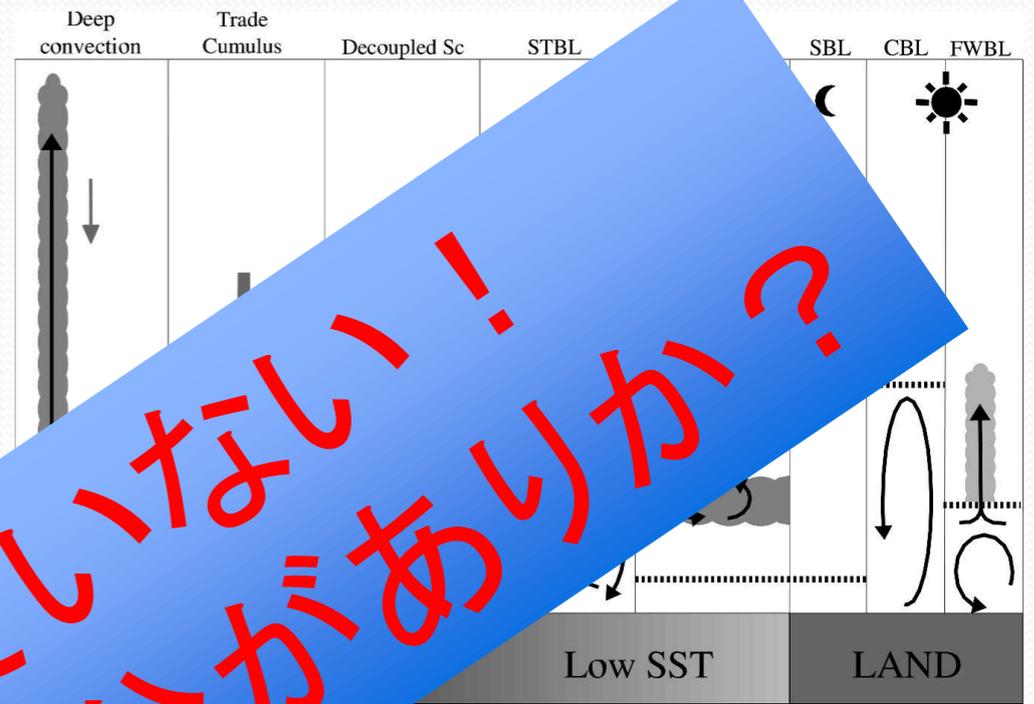
三浦



層積雲崩壊！予想されている描像（下から？上から？）

境界層のデカップリング （下からの対流不安定が制御）

- 熱帯へかけての海面水温上昇
 - 雲頂高度増加
- 境界層下層より新たな混合層の発現
 - 安定成層の2重化（混合層2重化）
 - 下層側の混合層：
水蒸気をため込む
鉛直不安定のポテンシャルをため込む
- 鉛直不安定解消のため積雲の崩壊

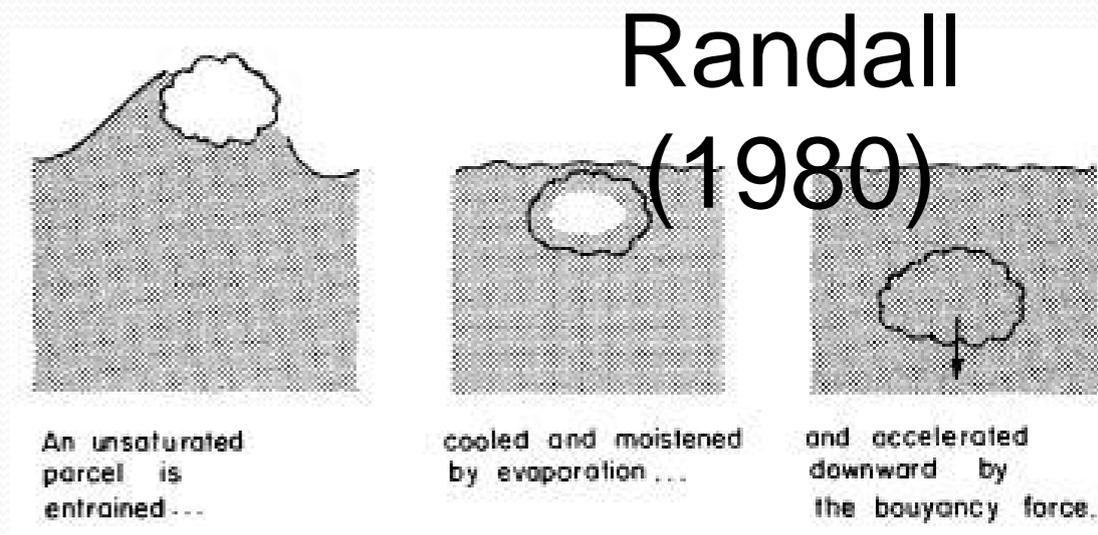


Stevens et al. (2007)

Cloud Top Entrainment I （上からの乾燥空気）

- 熱帯にかけての海面水温上昇
 - 雲頂高度増加
- 境界層下層より新たな混合層の発現
 - 安定成層の2重化（混合層2重化）
 - 下層側の混合層：
水蒸気をため込む
鉛直不安定のポテンシャルをため込む
- 鉛直不安定解消のため積雲の崩壊

決着は未だついていない！
デカップリングにやや分がありがたか？



層積雲崩壊と積雲発生は、今後の気候モデルの鍵！

サイエンスターゲット：

- **層積雲内の細かな物理を解きつつ**
- **層積雲→積雲への遷移過程を表す大規模領域計算**
- 提案されている崩壊メカニズムとメソスケール循環の関係
 - デカップリング？CTEI？別のメカニズム？
 - 特に、大規模場からの重力波がどのように寄与しているか？
 - **戦略分野3気象・気候分野と強く連携**
- 必要条件
 - 大規模場と遷移領域との相互作用を表現できる領域であること。
 - 1000km X 200km X 4km(鉛直) を想定。
 - 既往研究：数10kmX数10kmX4km
 - 微細な物理およびより原理的な乱流計算であること
 - 格子間隔20m程度 を想定。
 - 既往研究：格子間隔60mぐらい
 - 手法は、LES (Large Eddy Simulation)

実験設定詳細(2, 3月時点、現時点では若干の変更)

- 領域: 1000kmX200kmX4km
- 積分時間1日
- $Dx = 20m$ / $Dt = 0.058s$
- Step数: 1468800
- 1格子1stepあたりのおよその計算量
 - 力学コア: 3000FLOP + 物理2000FLOP (1step換算)
- 総演算量: 734EFLOP
- 実効20%を想定 全系で2PFLOPS
- Elapse time : 100時間(松)
 - 参考: 30m格子の場合には: 20時間 (竹) / 40m格子の場合には: 6時間(梅)
- 出力: 他分野よりは、出力大きい
 - 0.025TB / 一変数
 - 11変数 -> 0.275TB
 - 1分毎のデータ、計1440回
 - **トータル: 400TB**

それで現状計算パフォーマンスはど
んなもの？

単体性能

八代尚 (AICS)

- ✱ $\Delta xyz=5m$, $\Delta t(dyn)=0.008sec$, $\Delta t(phy)=0.8sec$, $(k,i,j)=(1256,32,32)$
- ✱ 2.4sec分の積分
: かなりGrid数を増やして、高負荷にした設定

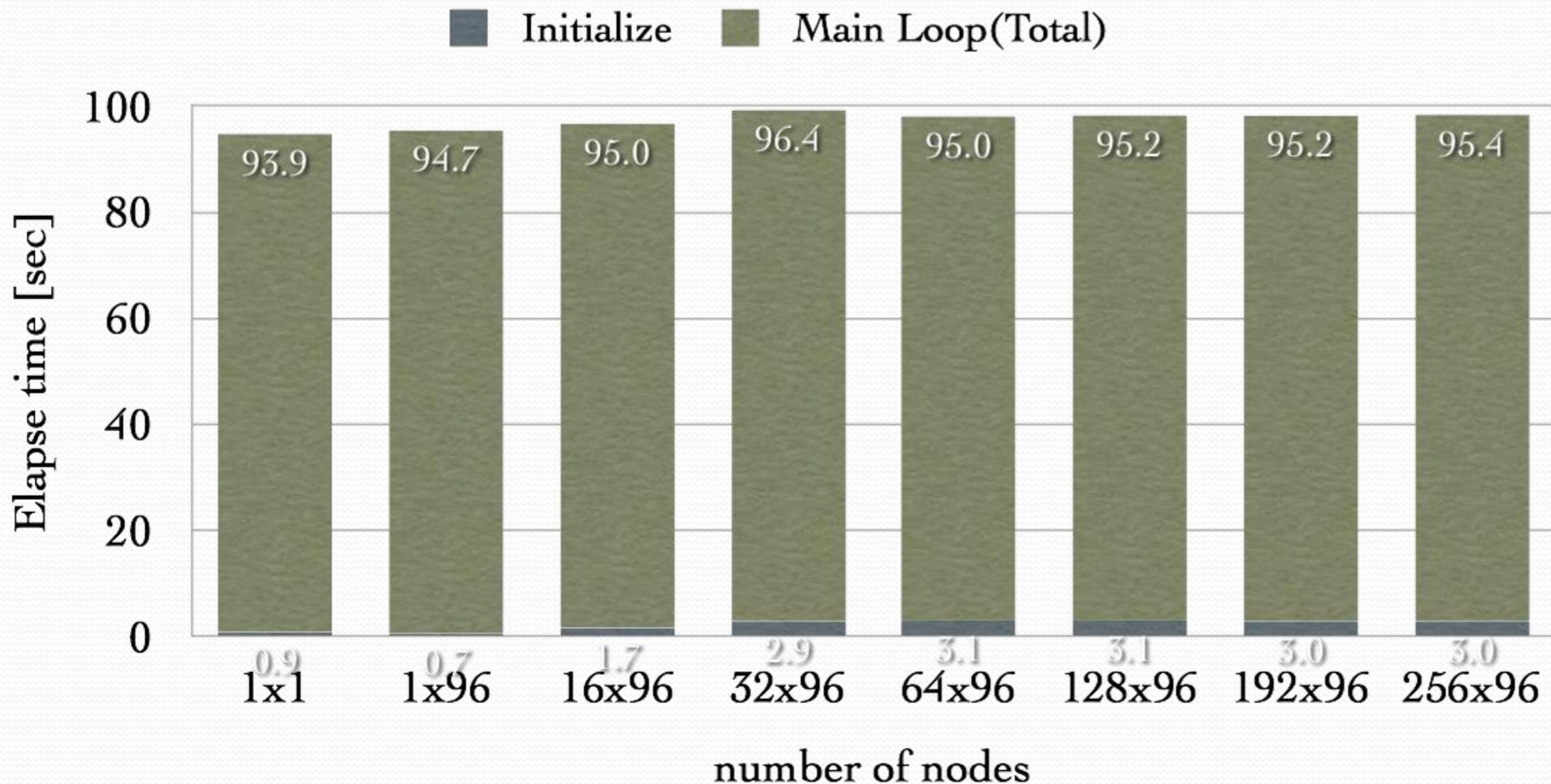
[sec]	CTL	RDMA	w/ kessler	w/ NDW6 orig	w/ NDW6 tuning
Main loop	110.6	102.5	112.7	124.9	116.3
Dynamics	109.5	101.5	108.7	108.7	108.7
TB	0.0	0.0	0.8	1.0	0.9
Micro-	0.0	0.0	1.6	13.7	5.4

力学分分だけだと、大体10%から12, 3%。
Real4にしたら、もうちょい上がる

スケーリング結果

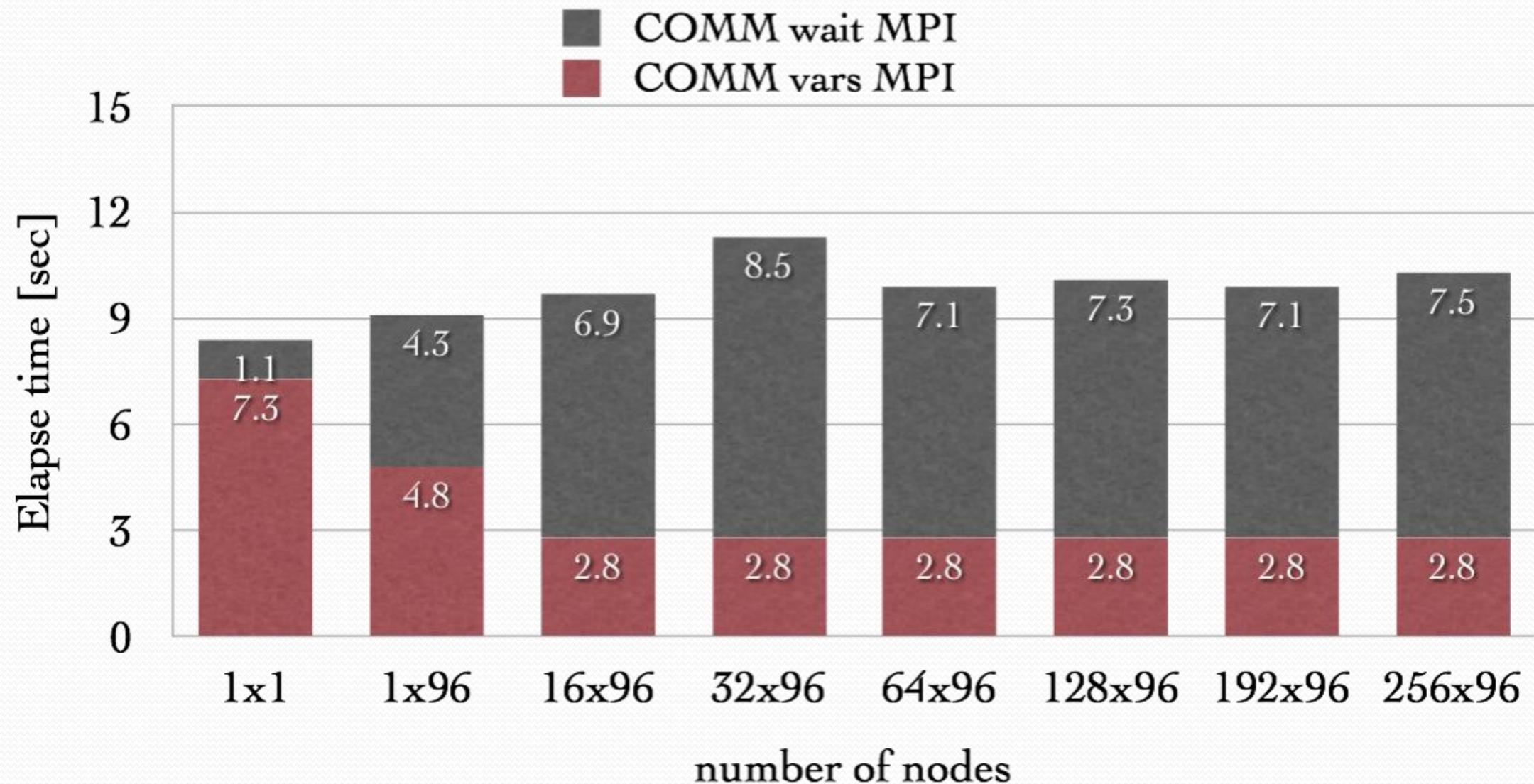
八代尚 (AICS)

- ✿ $\Delta xyz=5m$, $\Delta t(\text{dyn})=0.008\text{sec}$, $\Delta t(\text{phy})=0.8\text{sec}$, $(k,i,j)=(1256,32,32)$
- ✿ 2.4sec分の積分
: かなりGrid数を増やして、高負荷にした設定



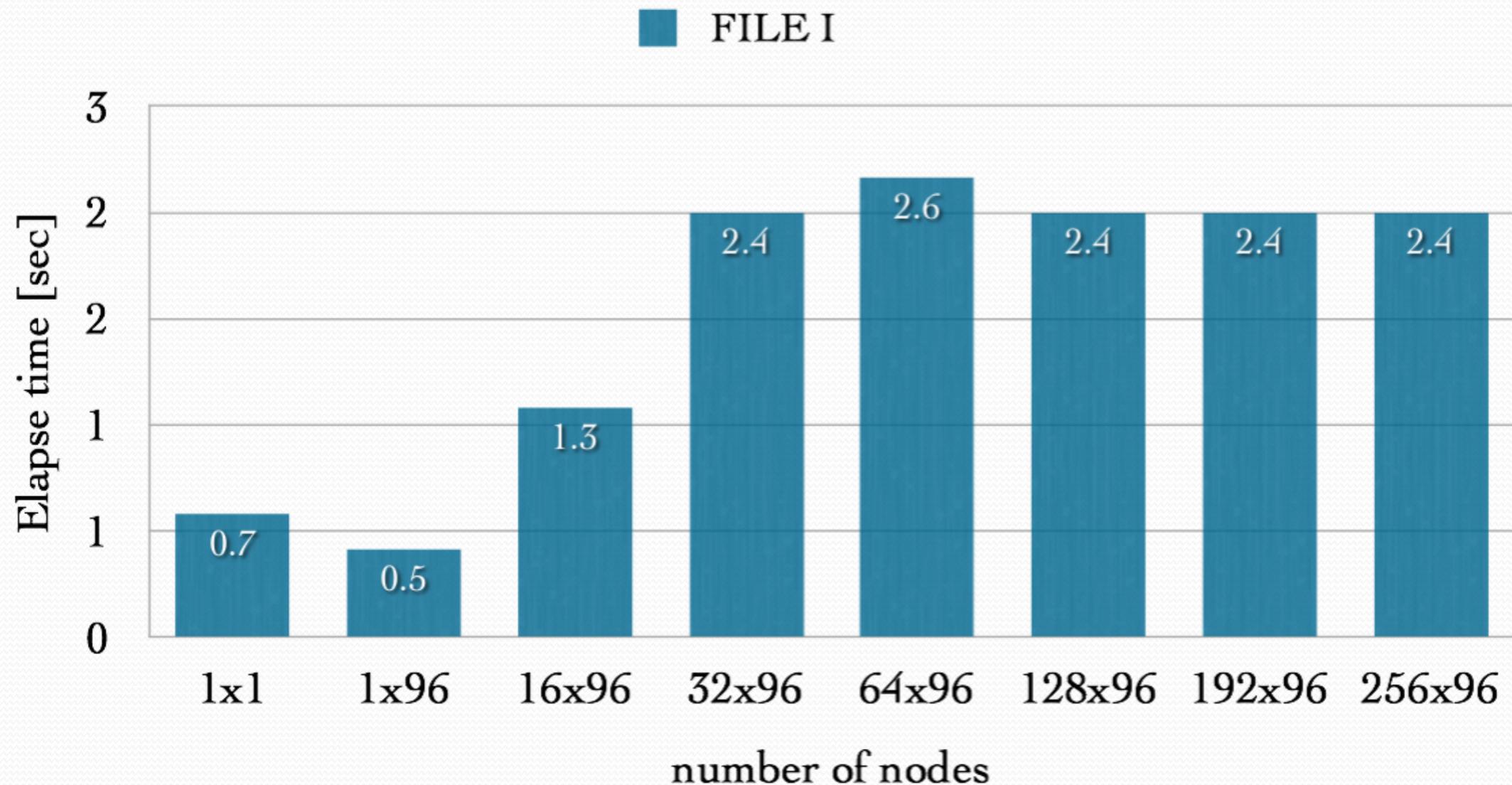
スケーリング結果

- ✿ $\Delta xyz=5m$, $\Delta t(\text{dyn})=0.008\text{sec}$, $\Delta t(\text{phy})=0.8\text{sec}$, $(k,i,j)=(1256,32,32)$
- ✿ 2.4sec分の積分
: かなりGrid数を増やして、高負荷にした設定



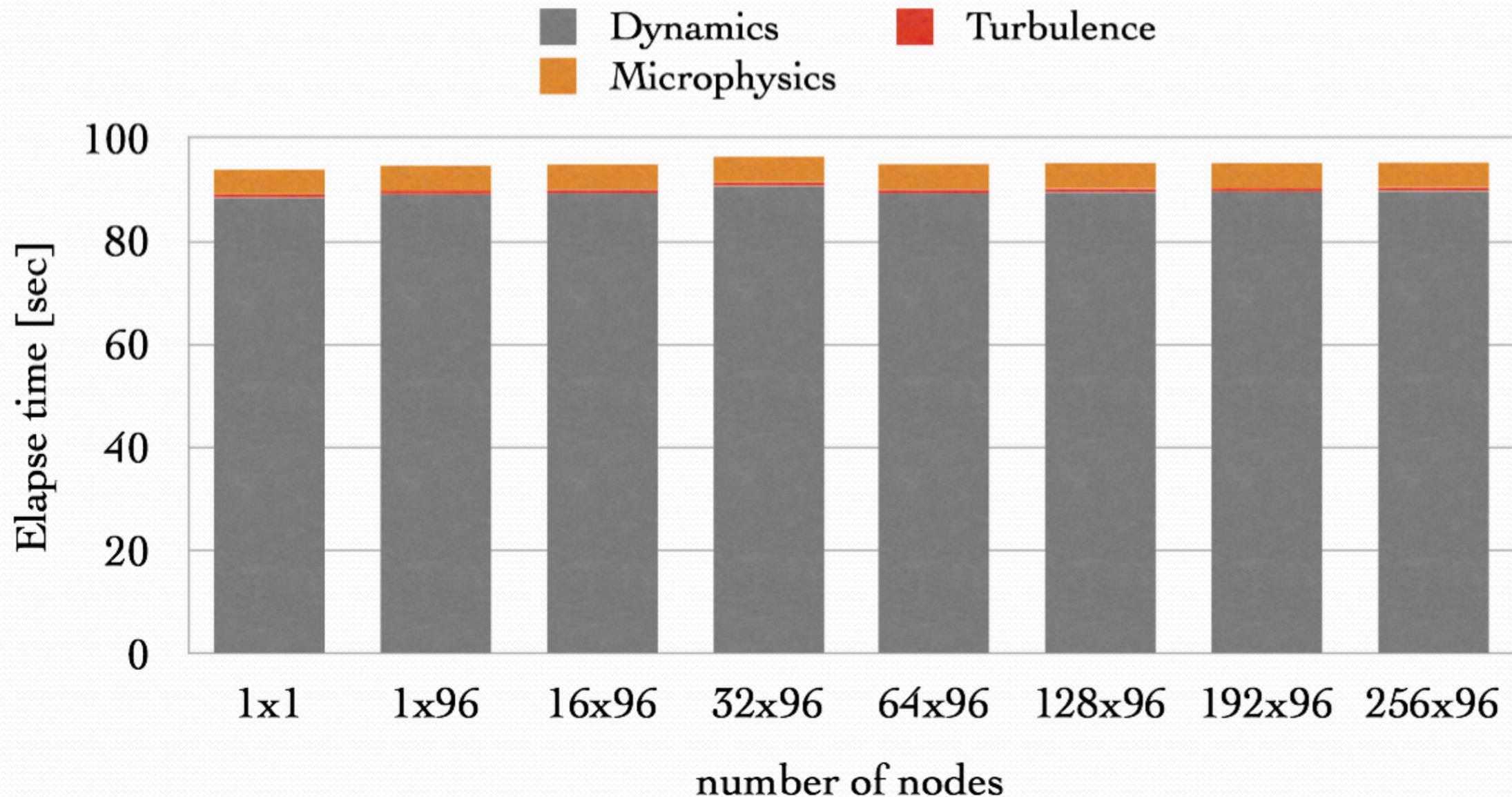
スケーリング結果

- ✿ $\Delta xyz=5m$, $\Delta t(\text{dyn})=0.008\text{sec}$, $\Delta t(\text{phy})=0.8\text{sec}$, $(k,i,j)=(1256,32,32)$
- ✿ 2.4sec分の積分
: かなりGrid数を増やして、高負荷にした設定



スケーリング結果

- ✿ $\Delta xyz=5m$, $\Delta t(\text{dyn})=0.008\text{sec}$, $\Delta t(\text{phy})=0.8\text{sec}$, $(k,i,j)=(1256,32,32)$
- ✿ 2.4sec分の積分
: かなりGrid数を増やして、高負荷にした設定



```
do large_step = 1, 3
  ! dynamics
  do small_step = 1, 100
    computation
    call COMM_rdma_vars8( 17 , 5 )
    computation
    call COMM_rdma_vars8( 22 , 5 )
    computation
    call COMM_rdma_vars8( 1 , 5 )
    computation
    do iq = 1, 11
      computation
      call COMM_rdma_vars8( 27 , 3 )
      computation
    enddo
    computation
    call COMM_rdma_vars8( 6 , 11 )
    computation
  enddo
enddo
```

- Dynamics 300 small step分の通信時間を測定
 - グリッドのサイズを, $(Z,X,Y) = (1256, 32+2*HALO, 32+2*HALO)$ とし, 1回の通信でHALOを隣接8方向(東西南北+斜め)に通信
 - 変数は29種類(29個の3次元配列)
 - 通信回数 4500回
 - 5配列の送受信 900回
 - 3配列の送受信 3300回
 - 11配列の送受信 300回
 - 今回の測定では, computationは空

COMM_rdma_vars8(vid , cnt)

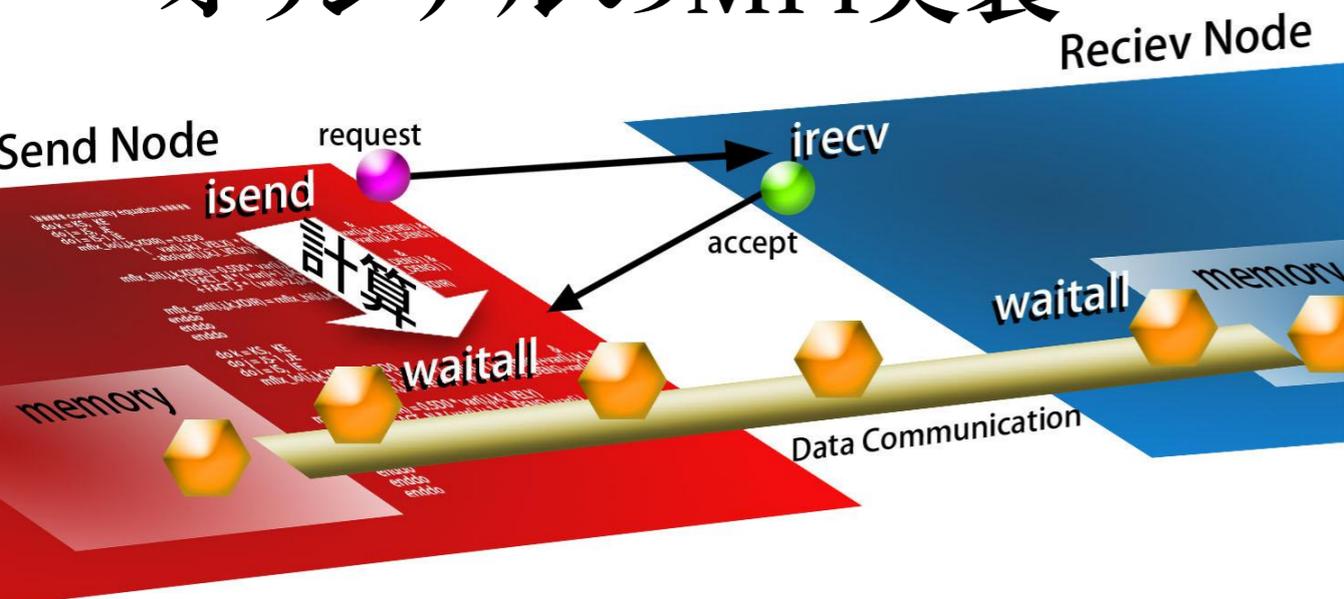
- 配列には integer の ID が割り振られており, ID を引数に与えて, 通信する配列を指示
- COMM_rdma_vars8 関数1回で, vid から vid+cnt-1 までの cnt個の配列を転送

MPI版では, 以下に置き換え

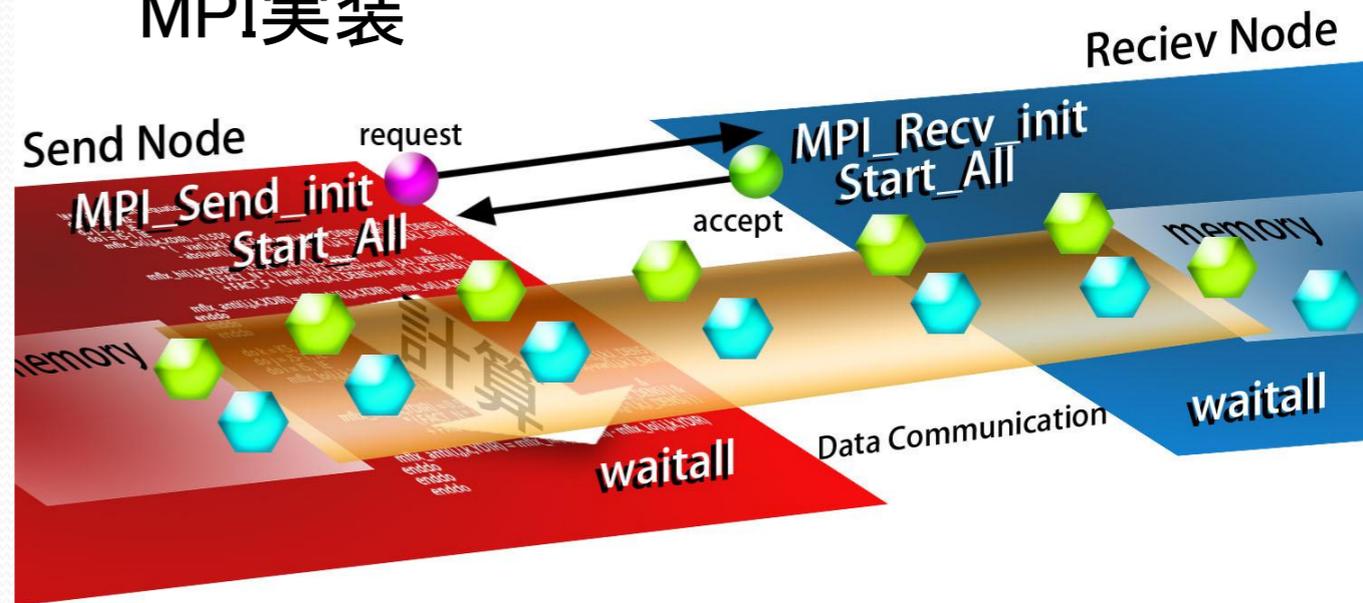
```
do i = 1, cnt
call COMM_vars8( var(:,:,:,vid+i-1) )
enddo
do i = 1, cnt
call COMM_wait( var(:,:,:,vid+i-1) )
enddo
```



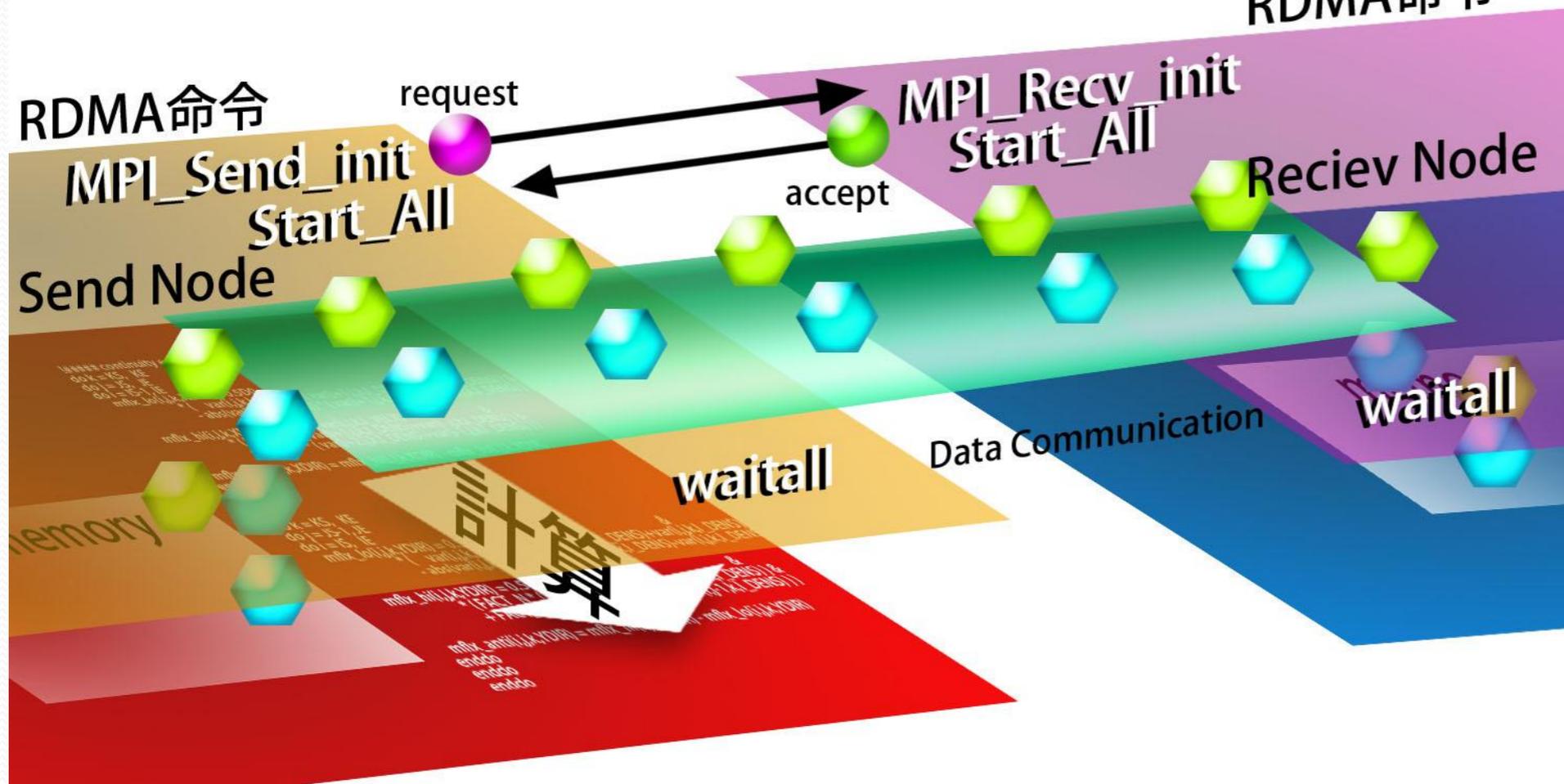
● オリジナルのMPI実装



● Persistent communicationを使ったMPI実装

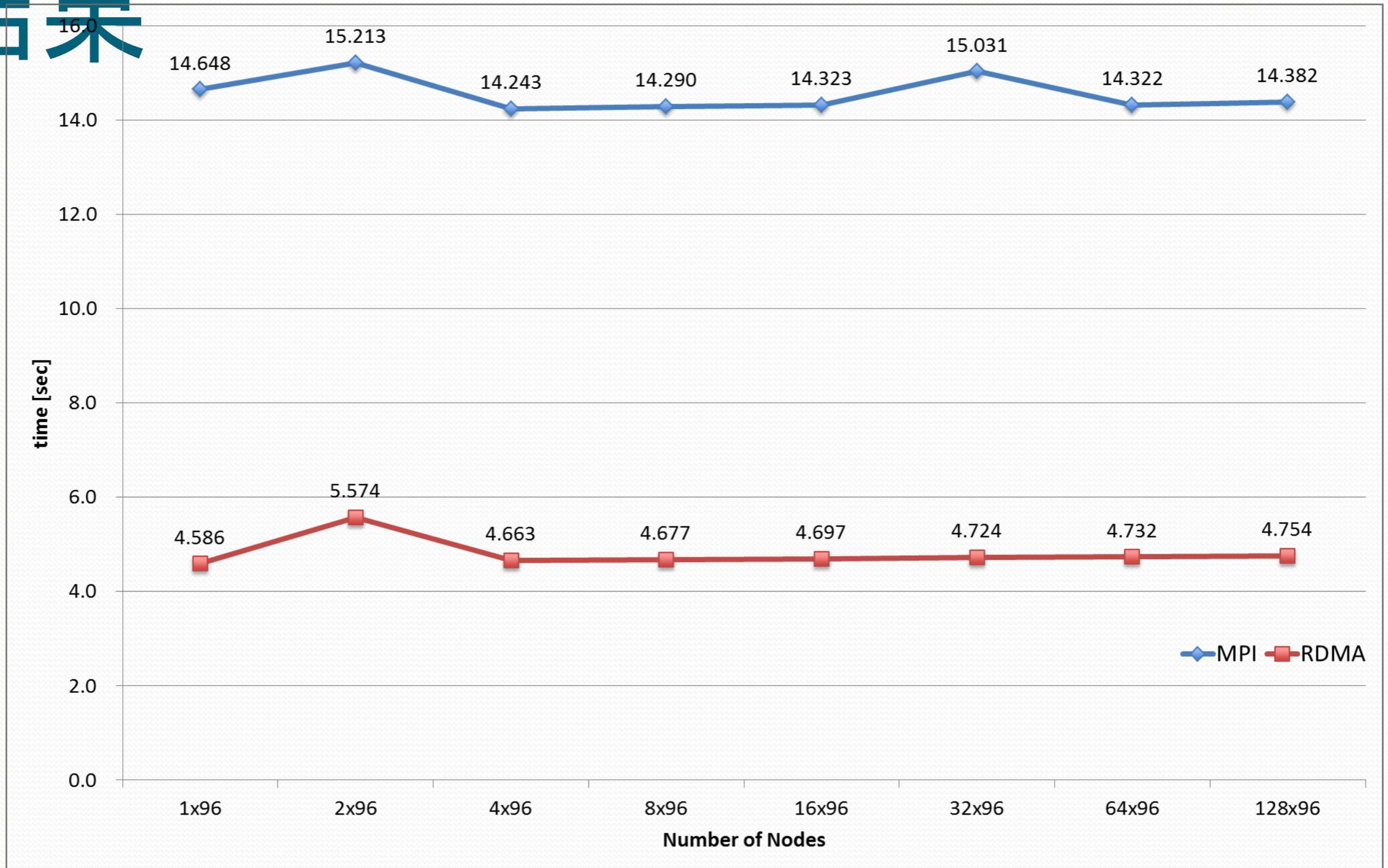


● さらなる改良: RDMAの直接操作(通信・計算オーバーラップ版) RDMA命令



- RDMAを使うと、MPIよりも3倍高速

結果



まとめ

- HPC動向をにらみながら、気象・気候・惑星科学統合ライブラリーを目指す。
 - 高速＋可読性のバランス
 - 共通のプラットフォームを作ること、基本的なインフラを整備する。
- まずは、「京」で広領域LESに挑戦。
 - テストケースはクリアしている。
 - モデルパフォーマンスは、単体性能、並列性能、まあ、イケてそうという感触。まだまだ、向上の余地あり。
 - 最初のサイエンスターゲット：層積雲の崩壊過程