



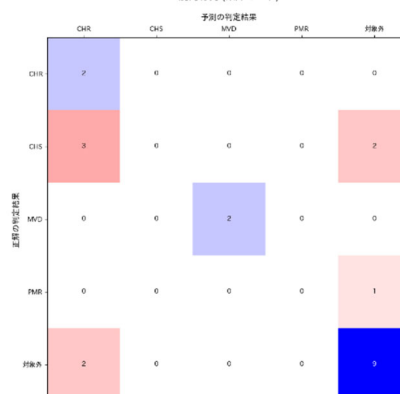
オンラインセミナー 生成 AI・大規模言語モデルと CAE セミナー

2023 年 9 月 22 日 (金) 開催

プログラム

1.	アドバンスソフト株式会社のご紹介 主催者あいさつ	1
招待講演①		
2.	「国内外のスパコンを用いた大規模事前学習の状況について」 東京工業大学 学術国際情報センター 先端研究部門 高性能計算先端応用分野 教授 横田 理央 様	5
招待講演②		
3.	「3次元形状生成 AI の産業応用に向けて」 名古屋大学 大学院工学研究科 土木工学専攻 構造・材料工学講座 准教授 西口 浩司 様	21
4.	AI 技術を活用した確率論的リスク評価手法の高度化研究 第 5 事業部 主管研究員 森本 達也	59
5.	生成 AI 導入・構築支援サービスについて 第 5 事業部 主管研究員 佐藤 智昭	73

混同行列 (故障モード)

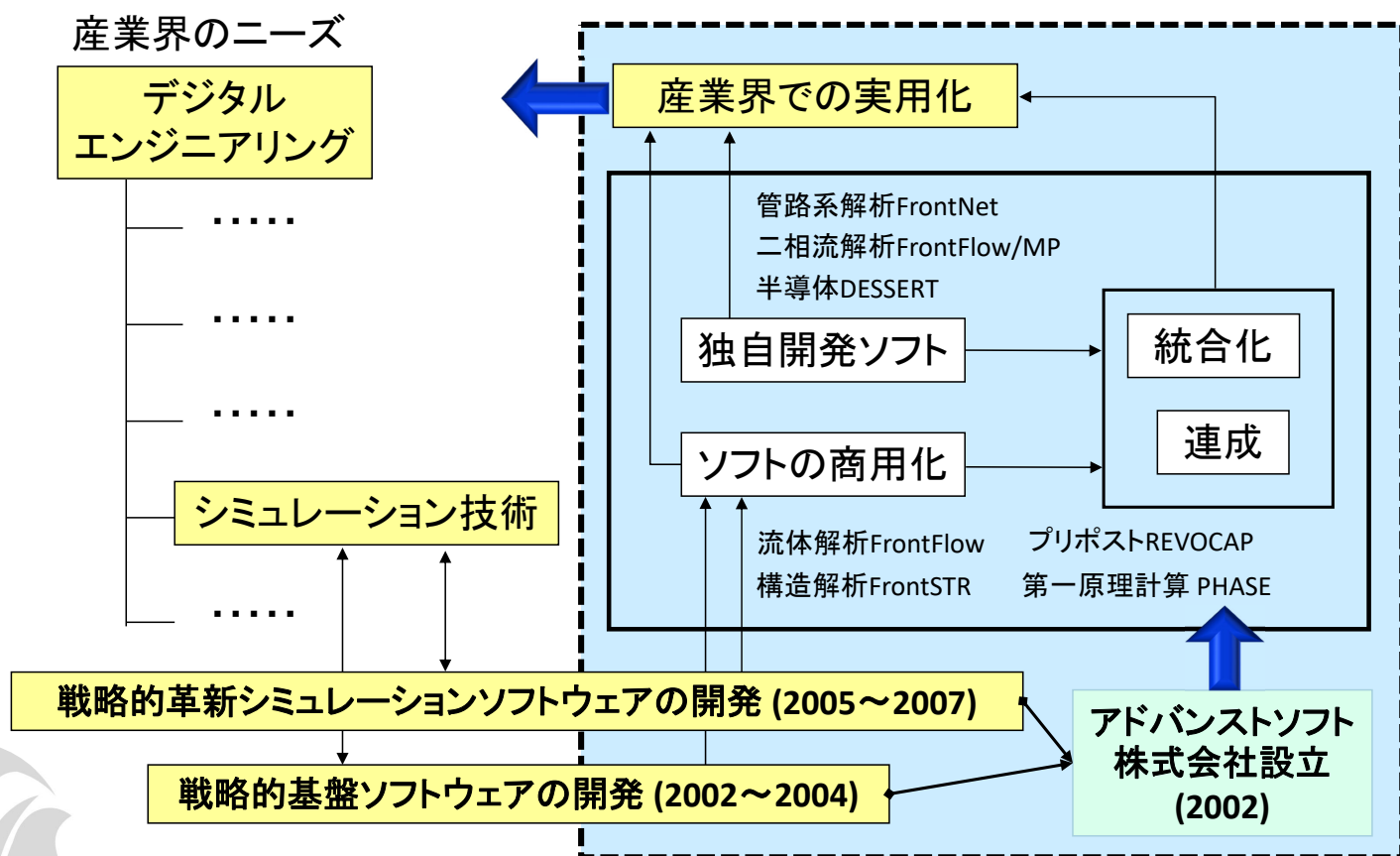


アドバンスソフト株式会社のご紹介

生成 AI・大規模言語モデルと CAE セミナー

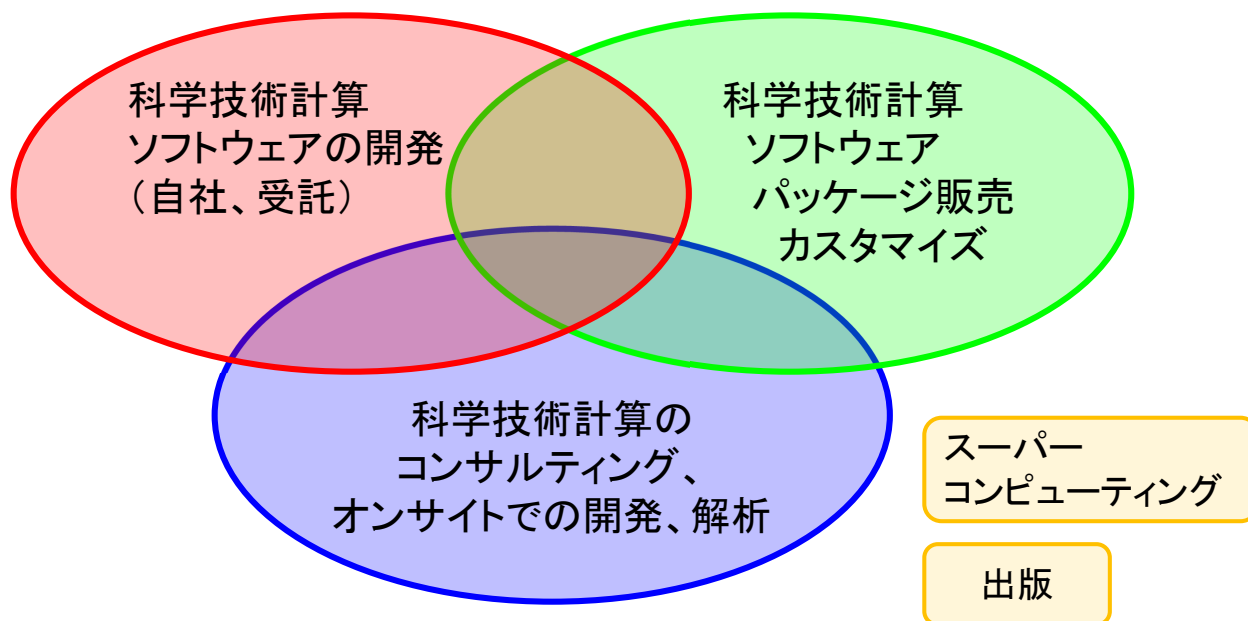
2023年9月22日（金）開催
アドバンスソフト株式会社

アドバンスソフトとは



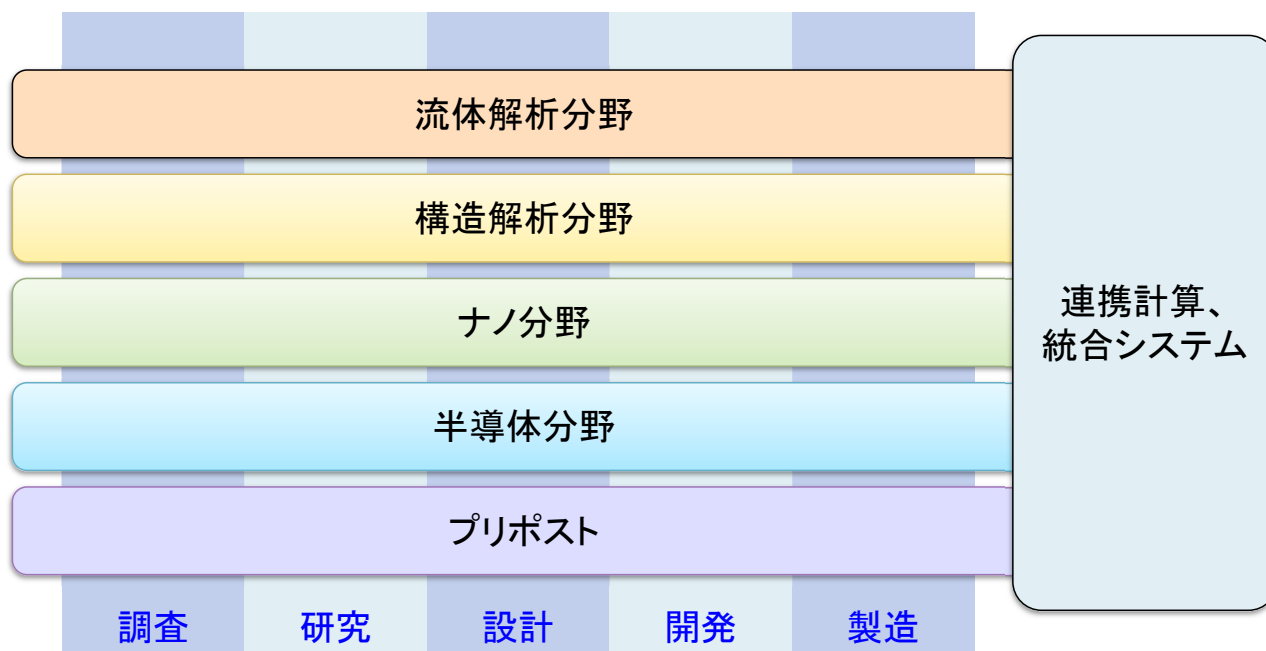
事業内容

アドバンスソフトがご提供するサービス



科学技術計算ソフトウェアの開発を基礎とした、
科学技術計算に関する様々なソリューションをご提供します。

事業分野



産業の主要な分野のあらゆるフェーズで直面する課題に対し、
科学技術計算によるソリューションをご提供します。

ソフトウェアご紹介

<p>第一原理計算ソフトウェア Advance/PHASE</p> <p>密度汎関数理論に基づき、物質の性質を原子・分子レベルから解析する第一原理計算ソフトウェアです。</p>	<p>ナノ材料 GUI 付属</p> <p>ナノ材料解析統合 GUI Advance/NanoLabo</p> <p>ナノ材料 プリポスト</p> <p>材料解析ソフトウェア QuantumESPRESSO と LAMMPS に対応した総合 GUI です。</p>	<p>流体解析ソフトウェア Advance/FrontFlow/red</p> <p>非圧縮性から圧縮性流れまで、広範囲で複雑な流れに対応した汎用 3 次元流体解析ソフトウェアです。</p>	<p>圧縮性流体解析ソルバー Advance/FOCUS-i</p> <p>非構造格子に対応した圧縮性流体解析ソルバーです。特に遷音速や超音速の流れに適しており、高い並列化効率で計算出来ます。</p>
<p>大規模 3 次元 TCAD システム Advance/TCAD</p> <p>超微細半導体デバイスからパワーデバイスまで、高度な機能と使いやすい GUI を備えた 3 次元 TCAD システムです。</p>	<p>半導体デバイス GUI 付属</p> <p>ニューラルネットワーク分子動力学システム Advance/NeuralMD</p> <p>ナノ材料 AI・機械学習</p> <p>Neural Network Potential に基づいた分子動力学のソフトウェアです。第一原理計算の結果を教師データとして分子力場を作成します。</p>	<p>気液二相流解析ソフトウェア Advance/FrontFlow/MP</p> <p>沸騰と凝縮を伴う気液二相流の流動特性や伝熱特性を 3 次元で解析するソフトウェアです。</p>	<p>管路系流体過渡解析ソフトウェア Advance/FrontNet</p> <p>配管や流体機器から成る管路系内流体に対する 1 次元過渡解析の実用的なソフトウェアです。</p>
<p>大規模電磁波解析ソフトウェア Advance/ParallelWave</p> <p>マクスウェル方程式を FDTD 法で 3 次元的に解く電磁波解析ソフトウェアです。アンテナの電波解析から光の干渉や回折を考慮した光波解析まで幅広く適用できます。</p>	<p>光波・電磁波</p> <p>構造解析ソフトウェア Advance/FrontSTR</p> <p>構造</p> <p>固体の変形や熱伝導を、有限要素法を用いた 3 次元で解析するソフトウェアです。</p>	<p>大気拡散影響予測システム Advance/Emerg</p> <p>GUI 付属</p> <p>大気拡散物質の挙動予測と影響評価のためのソフトウェアシステムです。</p>	<p>AI・機械学習</p> <p>深層学習用ツール Advance/iMacLe</p> <p>機械学習のうち、ニューラルネットワークによる深層学習に特化、最小限度の機能に絞り込んだ比較的軽いツールです。</p>
<p>汎用プリポストプロセッサ Advance/REVOCAP</p> <p>プリポスト</p> <p>解析の一連の流れをスムーズに行う事を実現した汎用プリポストプロセッサです。</p>	<p>音響解析ソフトウェア Advance/FrontNoise</p> <p>音響</p> <p>環境騒音、機器内の共振等における音場を有限要素法を用いた 3 次元で解析するソフトウェアです。</p>	<p>自社による開発 (国プロ含む) 開発チームによる質の高いサポートサービス カスタマイズや機能追加も応相談 並列数無制限 (追加料金なし)</p>	

ソフトウェアの解析事例

解析事例Webページをご覧ください。

アドバンスソフト 事例集

検索

<http://case.advancesoft.jp>

- ソフトウェア名からだけでなく、産業分野別、解析分野別の検索が可能となりました。
- 最新の事例を掲載しました。今後も逐次最新事例を紹介します。

産業分野別	解析分野別
自動車・運輸	流体
材料・化学	爆発・燃焼
産業機械	構造
航空宇宙	振動音響
エレクトロニクス	ナノ・バイオ
建設土木	プリポスト
原子力	半導体デバイス
エネルギー	光・電磁波
環境・防災	

facebook、YouTubeでも関連記事を掲載中

<http://www.facebook.com/advancesoft.jp>

<http://www.youtube.com/user/advancesoft>





国内外のスパコンを用いた 大規模事前学習の状況について

アドバンスソフト

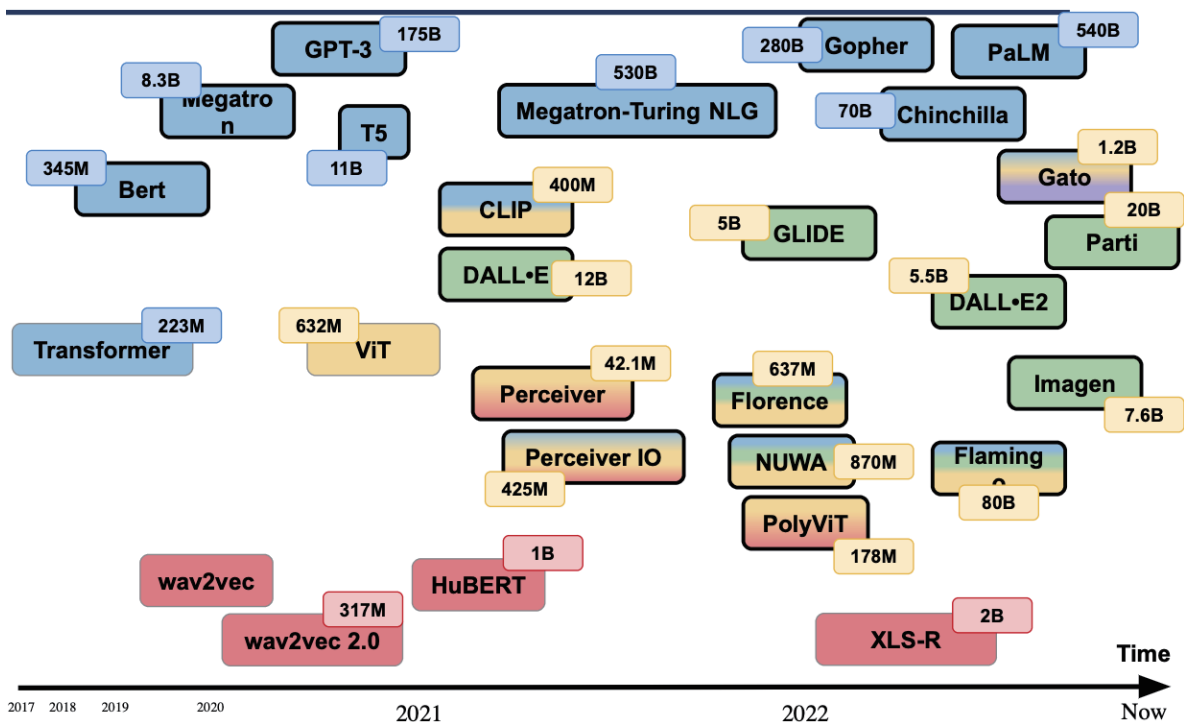
「生成 AI・大規模言語モデルと CAE」 セミナー

2023/9/22

Tokyo Institute of Technology
Rio Yokota
rioyokota@gsic.titech.ac.jp

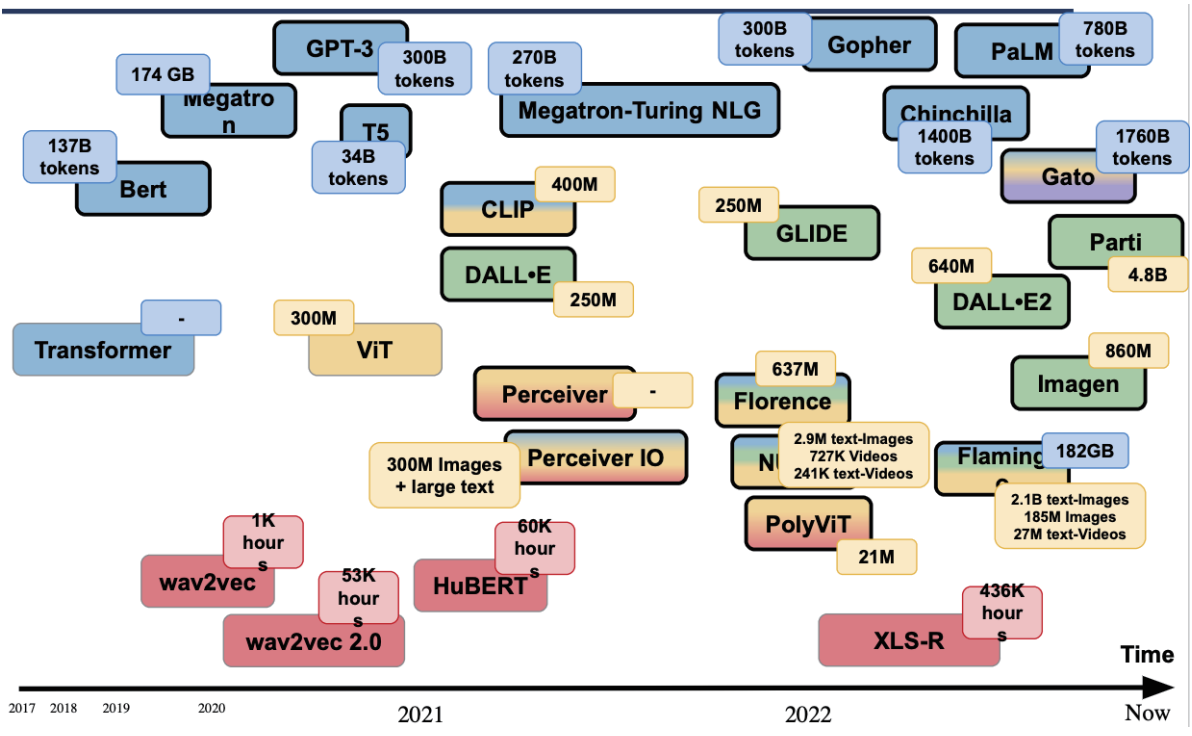
1

大規模深層学習 (パラメータ数)

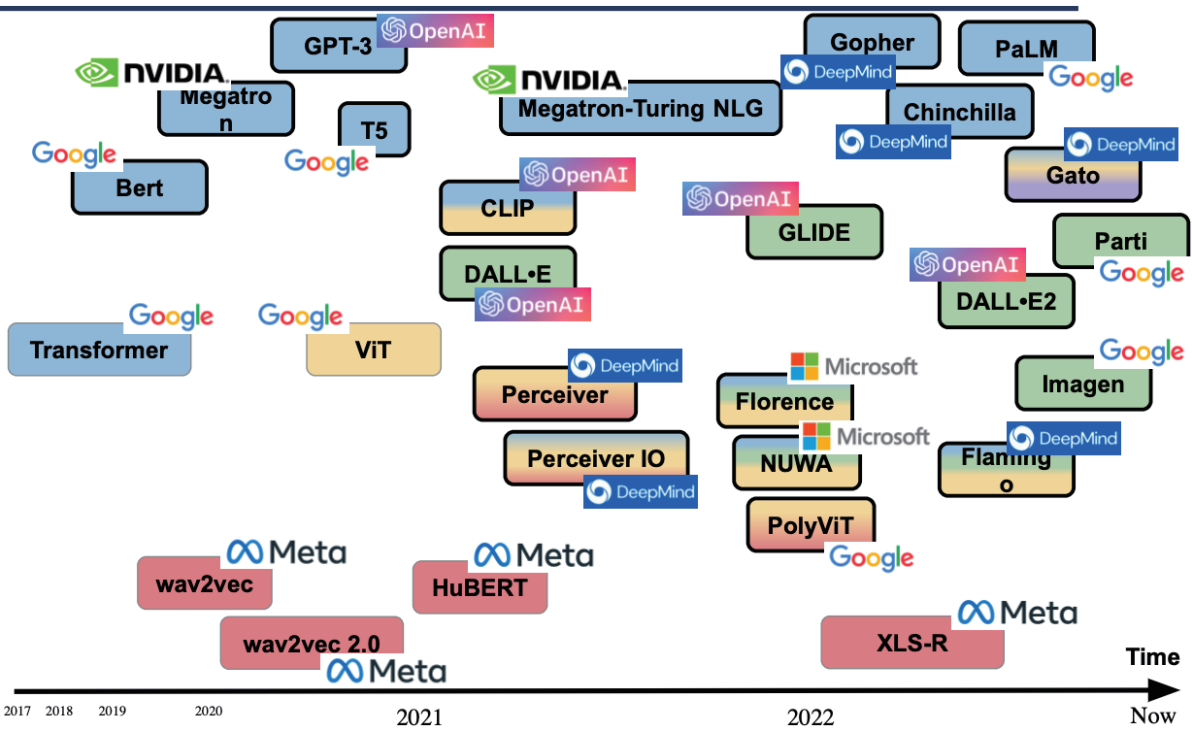


2

大規模深層学習 (データ量)



大規模深層学習 (所属)



- Transformerの性能がたった3つの変数のべき乗則に支配されている
- 現時点でScaling Lawの有効範囲には上限が見当たらず、性能に限界が存在しない可能性がある

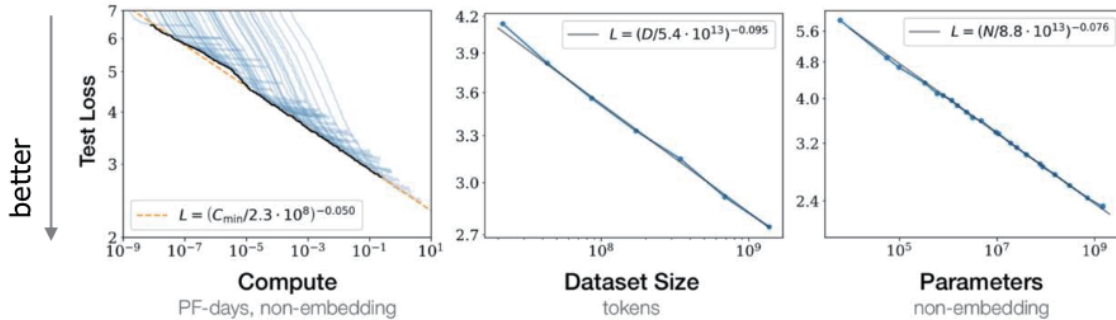
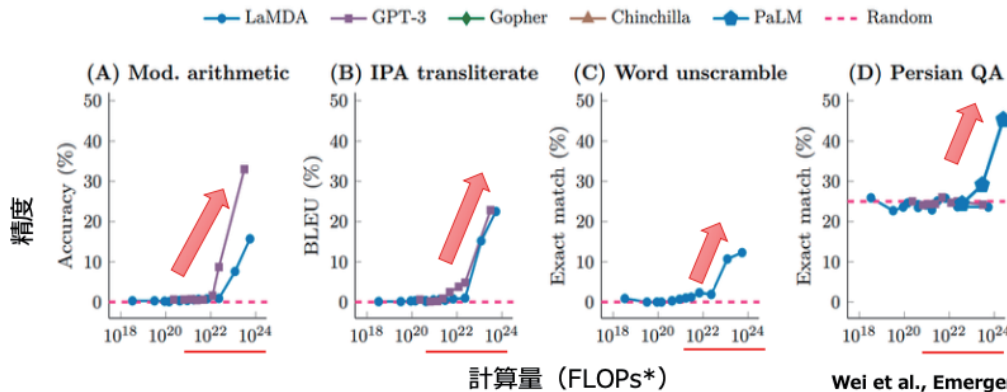


Figure 1 Language modeling performance improves smoothly as we increase the model size, dataset size, and amount of compute² used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

Scaling Laws for Neural Language Models
<https://arxiv.org/abs/2001.08361>

基盤モデルの創発性

- 基盤モデルで**10²³ FLOPs***の計算量の学習を行うと創発性を観測
 - それまで全くできなかったことが教えてもいないのに突然できるようになる
 - データ量と計算量だけ増やせばこの先も様々な能力を自然に獲得すると考えられる
 - GPTだけでなく他のモデルにおいても同様の規模で創発性が観測されている



10²³ FLOPs*で
精度が大きく改善

Wei et al., Emergent Abilities of Large Language Models, 2022

* ここでは **FLOPs** とはアプリケーションの **総演算量 (not 速度)** を指す

海外LLMの動向



Attention Is All You Need

Big Tech

- Google: Bard, Gemini
- OpenAI: GPT
- Meta: LLAMA, OPT
- Anthropic: Claude

- | | | | |
|---|---|--|--|
| Ashish Vaswani*
Google Brain
avaswani@google.com | Noam Shazeer*
Google Brain
noam@google.com | Niki Parmar*
Google Research
nikip@google.com | Jakob Uszkoreit*
Google Research
usz@google.com |
| Llion Jones*
Google Research
llion@google.com | Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu | Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com | |
| Illia Polosukhin* ‡
illia.polosukhin@gmail.com | | | |

Startups

- EleutherAI: GPT-NeoX, Pythia
- TogetherAI: RedPajama
- Databricks: Dolly
- MosaicML: MPT
- StabilityAI: StableLM

Startups by Transformer paper authors

- Adept.AI: Ashish Vaswani, Niki Parmar (2021/11-)
- Character.AI: Noam Shazeer (2021/11-)
- Inceptiv: Jakob Uszkoreit (2021/7-)
- Sakana.AI: Llion Jones (2023/8-)
- Cohere: Aidan N. Gomez (2019/9-)
- NEAR: Illia Polosukhin (2017/6-)

Universities

- Stanford: Alpaca
- Tsinghua: GLM

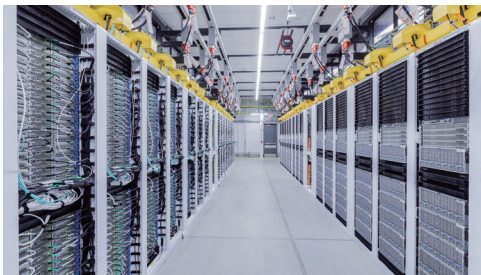
Huggingface Leaderboard

https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Top500 Supercomputer



OpenAI (25,000 GPU?)



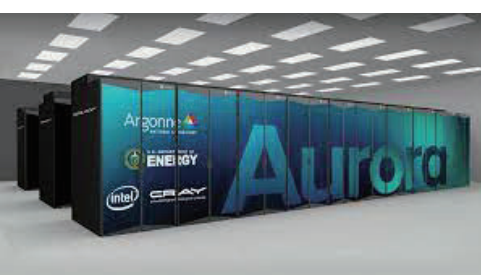
Frontier (37,888 GPU)



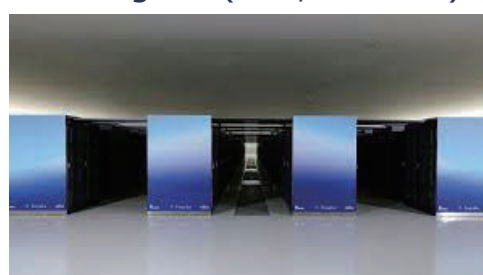
LUMI (20,480 GPU)



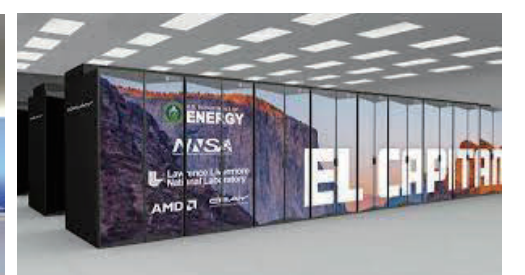
Aurora (63,744 GPU)



Fugaku (158,976 CPU)



El Capitan (20,000+ GPU)



INCITE Project

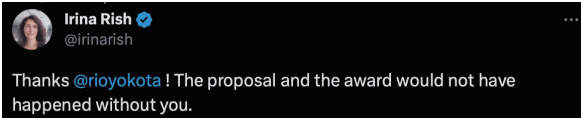
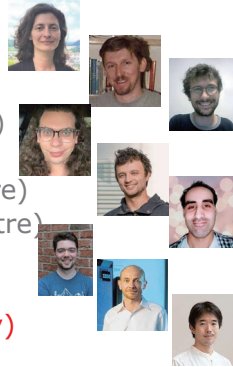


Type: New
Title: "Scalable Foundational Models for Transferable Generalist AI"

Principal Investigator: Irina Rish, University of Montreal, Mila - Quebec AI Institute
6M GPU hours on Summit

● **Project Members**

- Irina Rish (University of Montreal)
- Sergey Panitkin (University of Montreal)
- Guillaume Dumas (University of Montreal)
- Stella Biderman (EleutherAI)
- Jenia Jitsev (Jülich Supercomputing Centre)
- Mehdi Cherti (Jülich Supercomputing Centre)
- Quentin Anthony (Ohio State University)
- Guillermo Cecchi (IBM Research)
- **Rio Yokota (Tokyo Institute of Technology)**



	RedPajama	LLaMA*
CommonCrawl	878 billion	852 billion
C4	175 billion	190 billion
Github	59 billion	100 billion
Books	26 billion	25 billion
ArXiv	28 billion	33 billion
Wikipedia	24 billion	25 billion
StackExchange	20 billion	27 billion
Total	1.2 trillion	1.25 trillion



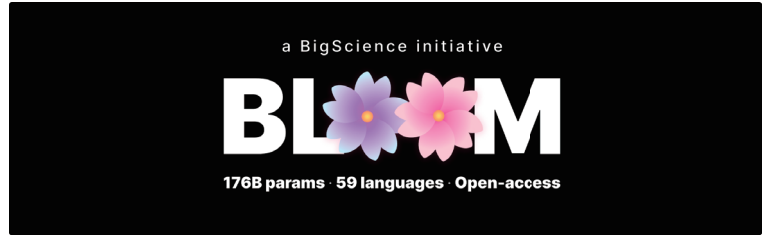
Aurora GenAI Project



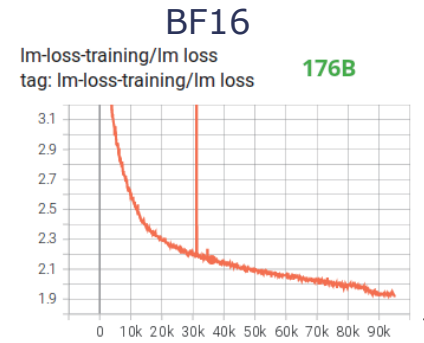
- Data: General text
 - Neeraj Kumar (PNNL) & Andrew McNaughton (PNNL)
- Data: Biological / medical
 - Arvind Ramanathan (Argonne) & Miguel Vazquez (BSC)
- Data: Chemistry / materials
 - Eliu Huerta (Argonne) & Gihan Pinipitiya (PNNL)
- Data: Physics
 - Salman Habib (Argonne), Paolo Calafiura (LBL)
- Data: Climate / environment
 - Po-Lun Ma (PNNL)
- Models: Evaluation, alignment, safety, and ethics
 - Bo Li (UIUC) & Prasanna Balaprakash (ORNL)
- Models: Downstream instruct tuning
 - Venkat Vishwanath (Argonne) & Väinö Hatanpää (CSC)
- Models: Pretraining runtime mixing monitoring
 - Shantenu Jha (BNL) & Juan Durillo (Liebniz LRZ)
- Models: Inference / optimization & architecture / performance
 - **Rio Yokota (Tokyo Tech.)** & Jeyan Thiyagalingam (RAL)
- AI for Computer Science
 - Valerie Taylor (Argonne) & Pete Beckman (Argonne)
- TPC coordination, strategies, and policy
 - Charlie Catlett (Argonne) & David Martin (Argonne)

BLOOM Project

Hardware	384 80GB A100 GPUs
Software	Megatron-DeepSpeed
Architecture	GPT3 w/ extras
Dataset	350B tokens of 59 Languages
Training time	3.5 months



- HuggingFace + IDRIS / GENCI
- Over 1000 AI researchers involved
- Trained on Jean Zay (France)
- 2022/3 - 2022/7
- 46 natural languages
- 13 programming languages



<https://huggingface.co/blog/bloom-megatron-deepspeed#datasets>

11

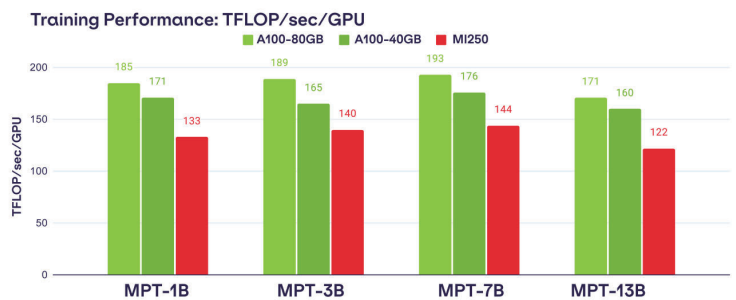
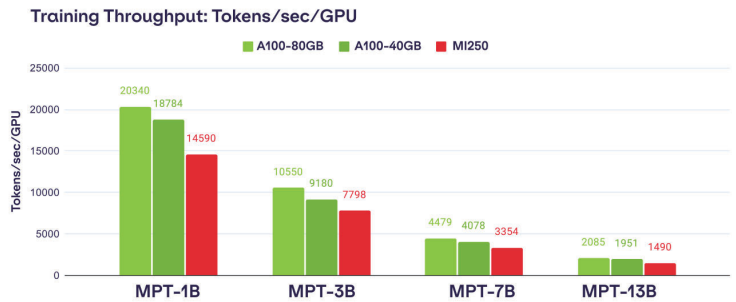
Projects on LUMI



Training LLMs on AMD GPUs



- University of Turku: TurkuNLP (13B)
- Allen Institute: AI2 OLMo (70B)
- LAION: Multilingual CLIP
- EuroHPC: Multilingual StarCoder



12

国内LLMの動向



Industry

- Cyberagent
- ELYZA
- Line
- NICT
- PFN
- Rinna
- StabilityAI

Universities/Labs

- NII(LLM勉強会): MDX
- 理研(GPT-Fugaku): 「富岳」
- 産総研: ABCI
- 松尾研

Weights & Biases Leaderboard
<http://wandb.me/nejumi>

Rakuda
<https://yuzuai.jp/benchmark>

Stability AI Leaderboard
<https://github.com/Stability-AI/lm-evaluation-harness/tree/jp-stable>

外国製 ↑

	model_name	平均点 ↓	MARC-ja	JSTS-pearson	JNLI	JSQuAD-F1	JCommonsenseQ
17	gpt-4	0.9048	0.9781	0.8902	0.7654	0.9492	0.941
8	stabilityai/StableBeluga2	0.771	0.9655	0.6629	0.4306	0.9092	0.8865
18	gpt-3.5-turbo	0.7698	0.954	0.8354	0.62	0.8407	0.5987
7	stabilityai/StableBeluga-13B	0.682	0.9515	0.7243	0.3135	0.87	0.5505
11	mosaicml/mpt-30b-instruct	0.4469	0.8511	0.03051	0.145	0.8154	0.3923
20	mosaicml/mpt-7b-instruct	0.377	0.8495	0.0574	0.2399	0.545	0.193
21	rinna/japanese-gpt-neox-3.6b-instruction-ppo	0.356	0.9558	0.2661	0.145	0.2202	0.193
4	line-corporation/japanese-large-lm-3.6b	0.3553	0.8463	0.3874	0.145	0.2046	0.193
1	matsuo-lab/webllab-10b-instruction-sft	0.3433	0.853	-0.09396	0.145	0.6192	0.193
6	stabilityai/japanese-stablelm-instruct-alpha-7b	0.3315	0.7989	0.03619	0.145	0.4842	0.193
19	rinna/japanese-gpt-neox-3.6b-instruction-sft	0.3141	0.9625	0.1303	0.145	0.1407	0.1921
3	stockmark/gpt-neox-japanese-1.4b	0.2681	0.8544	-0.001339	0.145	0.1493	0.193
5	line-corporation/japanese-large-lm-1.7b	0.2679	0.8544	-0.04536	0.145	0.1922	0.193
2	meta-llama/Llama-2-7b-chat-hf	0.2652	0.8976	0	0.145	0.09017	0.193
10	Salesforce/xgen-7b-8k-inst	0.2416	0.8212	-0.07148	0.145	0.1201	0.193
12	cyberagent/open-calm-7b	0.2411	0.8546	-0.01269	0.145	0.03199	0.1868

LLM勉強会



主催者：黒橋（国立情報学研究所）

基盤センター：北海道大学情報基盤センター、東北大学サイバーサイエンスセンター、東京大学情報基盤センター、東京工業大学学術国際情報センター、名古屋大学情報基盤センター、京都大学学術情報メディアセンター、大阪大学サイバーメディアセンター、九州大学情報基盤研究開発センター

大学の研究室：東北大学乾研究室、東北大学鈴木研究室、東京大学今泉研究室、東京大学大関研究室、東京大学川原研究室、東京大学鶴岡研究室、東京大学松尾研究室、東京大学宮尾研究室、東京大学谷中研究室、東京大学吉永研究室、東京大学医療AI開発学講座、早稲田大学河原研究室、東京工業大学岡崎研究室、東京工業大学横田研究室、お茶の水女子大学小林研究室、名古屋大学武田・笹野研究室、京都大学黒橋研究室、大阪大学鬼塚研究室、奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室、愛媛大学人工知能研究室

研究所：理化学研究所AIP、理化学研究所GRP、産業技術総合研究所、国立情報学研究所、情報通信研究機構、科学技術振興機構、情報・システム研究機構、

企業：NTT、LINE/ヤフー、レトリバ、サイバーエージェント、富士通、Microsoft、Studio Ousia、プレジジョン、ZENKIGEN、Legalscape、Turing、AWS、みらい翻訳、Megagon Labs、ストックマーク、matsuri technologies、ファーストアカウンティング、東芝、Preferred Networks、オムロンサイニックエックス、トヨタ、NTT Communications、パオパブ、Polaris.AI、Stability AI Japan、マネーフォワード、メルカリ、NVIDIA、アステラス製薬株式会社、パソコ、朝日新聞社、楽天、ELYZA、ベルシステム24、Lightblue、Intel



計算資源：A100 x 128 x 60 days

フレームワーク：Megatron-DeepSpeed、GPT-Neox、LLM-Foundry

日本語データ：Wikipedia: 1.4Bトークン (1.3M文書)
 mC4 (ウェブコーパス): 136Bトークン (75M文書)
 Common Crawl全量: 1Tトークン? (1B文書?)
 JST J-STAGE (論文): 3Bトークン程度 (5.5M文書)
 NDL WARP (ウェブアーカイブ): 1Tトークン以上?

モデルサイズ：1.3B、7B、13B

データ並列

データは分散
モデルは冗長
勾配の集約通信

課題：バッチサイズの増大に伴う汎化性能の低下

解決策：正則化・最適化手法の工夫

テンソル並列

データは冗長
モデルは分散
活性の集約通信

課題：通信頻度の増加

解決策：通信のオーバーラップ

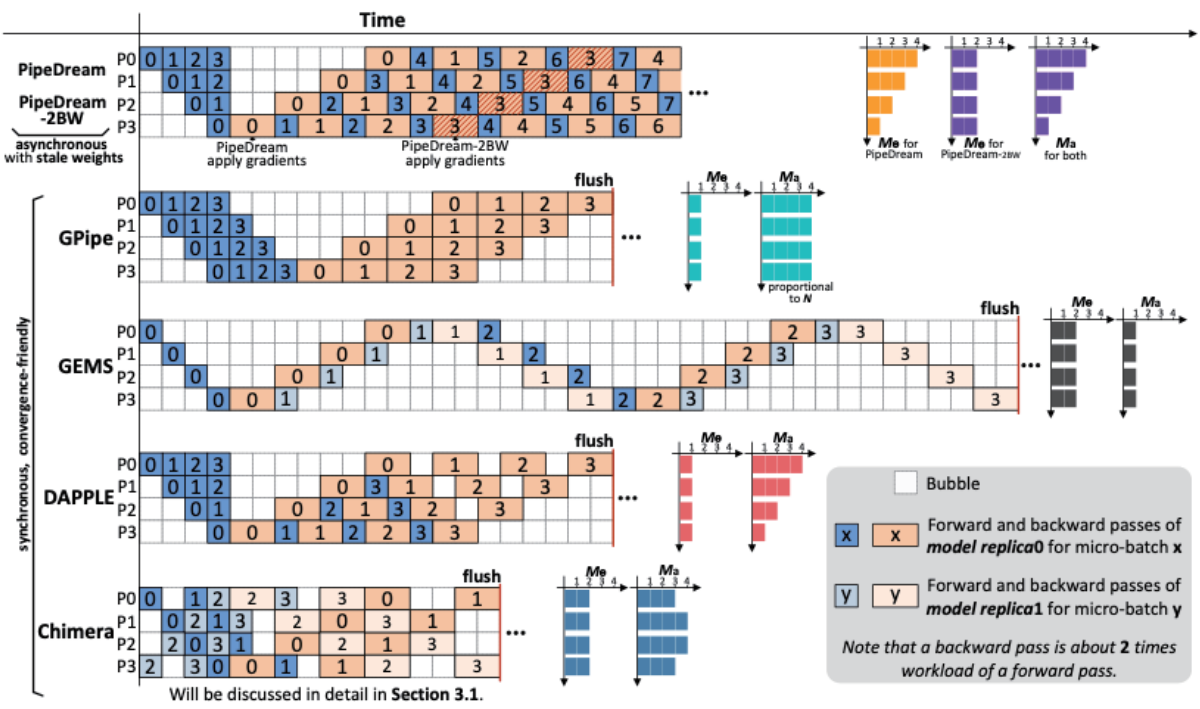
パイプライン並列

データは冗長
モデルは分散
活性の1対1通信

課題：パイプラインバブル

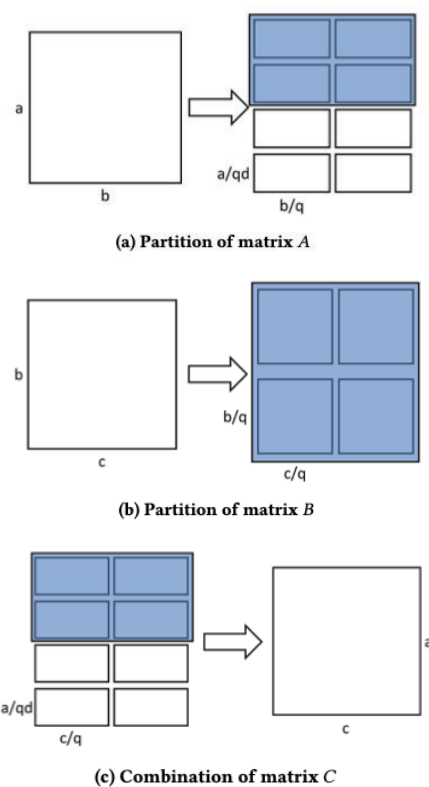
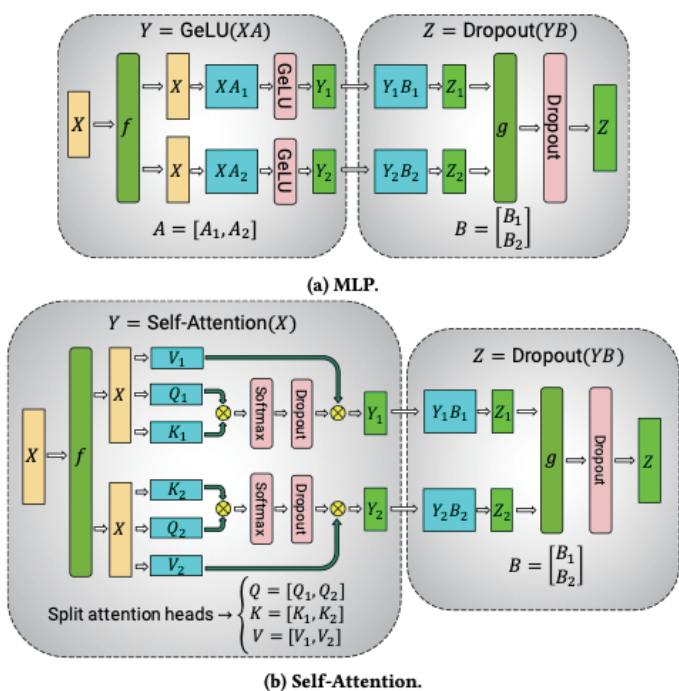
解決策：パイプラインの工夫

パイプライン並列



Chimera: Efficiently Training Large-Scale Neural Networks with Bidirectional Pipelines, <https://arxiv.org/abs/2107.06925>

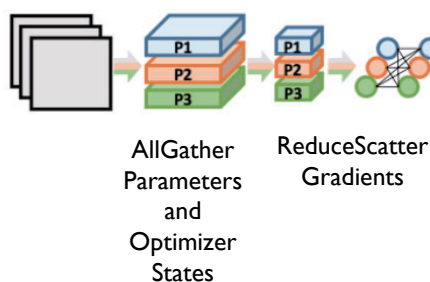
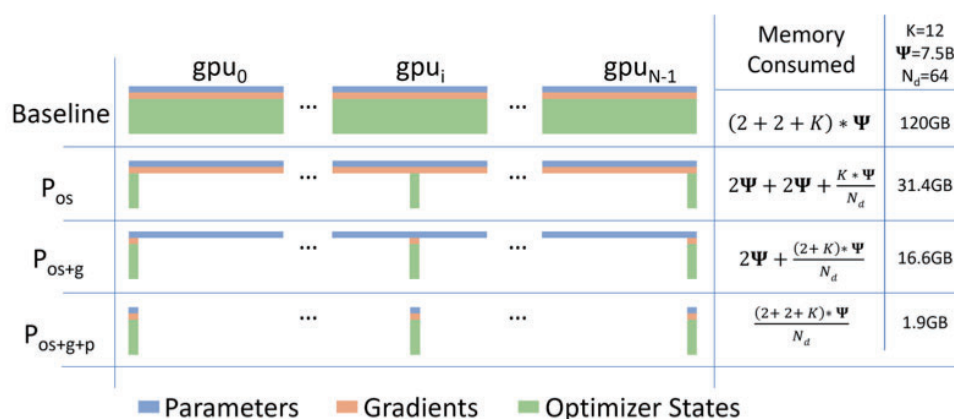
テンソル並列



Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM
<https://arxiv.org/abs/2104.04473>

Tesseract: Parallelize the Tensor Parallelism Efficiently
<https://arxiv.org/abs/2105.14500>

Zero Redundancy Optimizer (ZeRO)



Name	Optimizer + Grad (devices/partitioned)	Parameters (devices/partitioned)
Data parallel	[GPU] / ✗	[GPU] / ✗
ZeRO 2	[GPU] / ✓	[GPU] / ✗
ZeRO-Offload	[CPU,GPU] / ✓	[GPU] / ✗
3D Parallelism	[GPU] / ✓	[GPU] / ✓
ZeRO 3	[GPU] / ✓	[GPU] / ✓
ZeRO-Inf-CPU	[CPU, GPU] / ✓	[CPU,GPU] / ✓
ZeRO-Inf-NVMe	[NVMe,CPU,GPU] / ✓	[NVMe,CPU,GPU] / ✓

ZeRO-Infinity: Breaking the GPU Memory Wall for Extreme Scale Deep Learning, <https://arxiv.org/abs/2104.07857>

並列数の組み合わせ調査 (LLM勉強会)



フレームワーク: Megatron-DeepSpeed

Global batch size: 1536, 4096

Number of nodes: 8, 16

Data parallel: 16, 32

Pipeline parallel: 4, 8

Network: NCCL+RDMA, MPI+UCX

日本語データ: Wikipedia: 1.4Bトークン (1.3M文書)

mC4 (ウェブコーパス): 136Bトークン (75M文書)

Common Crawl全量: 1Tトークン? (1B文書?)

JST J-STAGE (論文): 3Bトークン程度 (5.5M文書)

NDL WARP (ウェブアーカイブ): 1Tトークン以上?

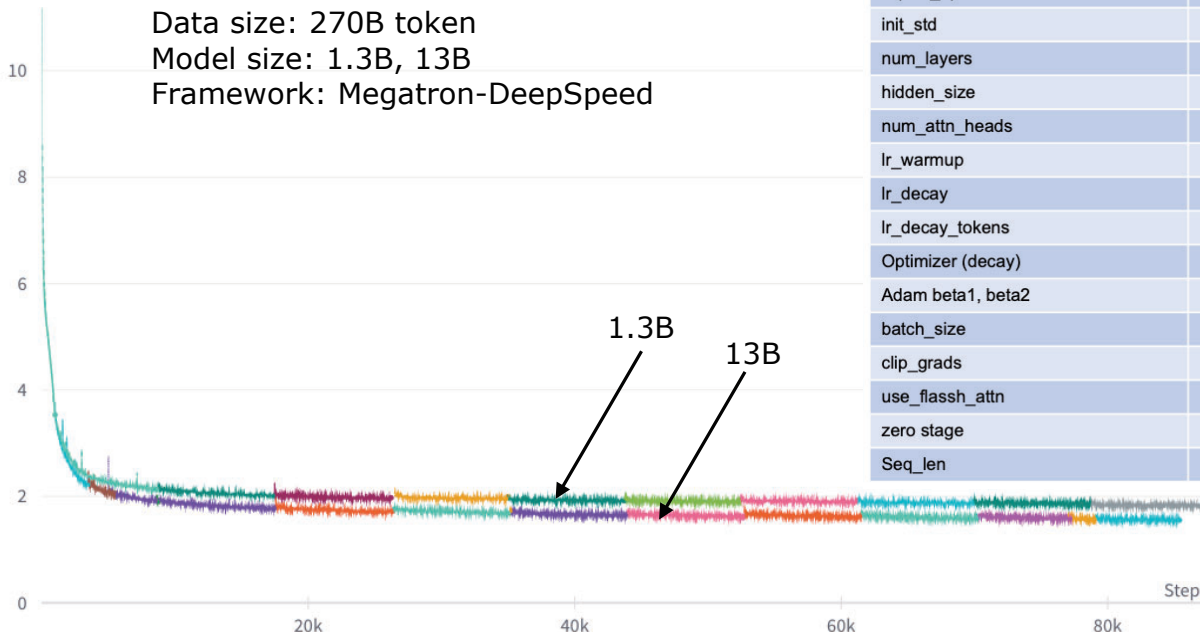
サイバーエージェント: 300B トークン

#nodes	GBS	NIC × ンテ	通信プロ トコル	DP, PP (TP=1)	TFLOPS / GPU	Tokens / sec	Days for 260B
16	1536	前	nccl + RDMA	16, 8	80.9	124K	24.2
				32, 4	42.8	66K	45.7
		後	MPI +UCX	16, 8	67.6	104K	29.0
				32, 4	49.4	76K	39.6
	4096	前	nccl + RDMA	16, 8	129.3	203K	14.8
				32, 4	133.2	209K	14.4
		後	MPI +UCX	16, 8	108.0	166K	18.1
				32, 4	114.2	176K	17.1
8	1536	前	nccl + RDMA	8, 8	143.2	110K	27.3
				16, 4	115.8	89K	33.8
		後	MPI +UCX	8, 8	143.1	110K	27.3
				16, 4	113.5	87K	34.5
	4096	前	nccl + RDMA	8, 8	143.4	110K	27.3
				16, 4	146.6	113K	26.7
		後	MPI +UCX	8, 8	147.5	113K	26.5
				16, 4	131.0	101K	29.9
後	MPI +UCX	8, 8	140.2	108K	27.9		
		16, 4	156.4	120K	25.0		

学習曲線 (LLM勉強会)



Data size: 270B token
Model size: 1.3B, 13B
Framework: Megatron-DeepSpeed



	1.3B	13B
lr (min_lr)	2.0e-4 (1.0e-6)	1.0e-4 (1.0e-6)
init_std	0.013	0.008
num_layers	24	40
hidden_size	2048	5120
num_attn_heads	16	40
lr_warmup	2000	2000
lr_decay	cosine	cosine
lr_decay_tokens	1080B	1080B
Optimizer (decay)	Adam (0.1)	Adam (0.1)
Adam beta1, beta2	0.9, 0.95	0.9, 0.95
batch_size	1536	1536
clip_grads	1.0	1.0
use_flash_attn	yes	yes
zero stage	1	1
Seq_len	2048	2048

GPT-Fugaku



メンバー：東工大、東北大、名古屋大、理研、富士通、サイバーエージェント、kotoba technologies

計算資源：「富岳」 約 5000万 ノード時間

フレームワーク：Megatron-DeepSpeed（富岳版）

モデルサイズ：1.3B、7B、13B、30B、175B

「富岳」でGPT-4を学習するには？

GPT-4: 3×10^{25} FLOPs (予測値)

GPT-3.5 (ChatGPT): 3×10^{24} FLOPs (予測値)

GPT-3: 3×10^{23} FLOPs

「富岳」：

FP32 6.76TFLOP/s x 158,976 = 1.07 EFLOP/s (理論ピーク性能)

GPT-4: 328 days x 4

GPT-3.5: 32 days x 4

GPT-3: 3.3 days x 4

OpenAI:

BF16 312 TFLOP/s x 25,000 = 7.8 EFLOP/s (理論ピーク性能)

GPT-4: 45 days x 2

GPT-3.5: 4.5 days x 2

GPT-3: 11 hours x 2

Googleは現在この5倍の資源でGeminiを学習中

大規模コーパス



コーパス	ソース	サイズ(GB)	補足説明
CyberAgent CommonCrawl (ja)	Common Crawl	1,300	サイバーエージェント社が独自に構築
mc4 (ja)	Common Crawl	97	
日本語Webコーパス	東北大学	63	東北大学が独自に構築
Oscar (ja)	Common Crawl	37	
Wiki-40B(ja)	日本語Wikipedia	2	
WMT22 (en-ja)	Multiple	9	対訳データ（翻訳能力+英語からの知識移転）
News (ja, en-ja)	毎日、朝日、日経、読売新聞	~1.5/年	商用利用には別途契約が必要（現在交渉中）
Twitter (ja)	東北大学	10~	言語モデルの事前学習には適さない（短文）

上記に加えて、Wikipedia (en), WikiHow (en, ja), Presto (en, ja), bigcode (program language) といった質が高く、かつ言語モデルのユースケース（対話やHowTo、QA）に近いデータも用いる。

「富岳」でGPTを学習するには？

○基盤モデルの事前学習の計算量：

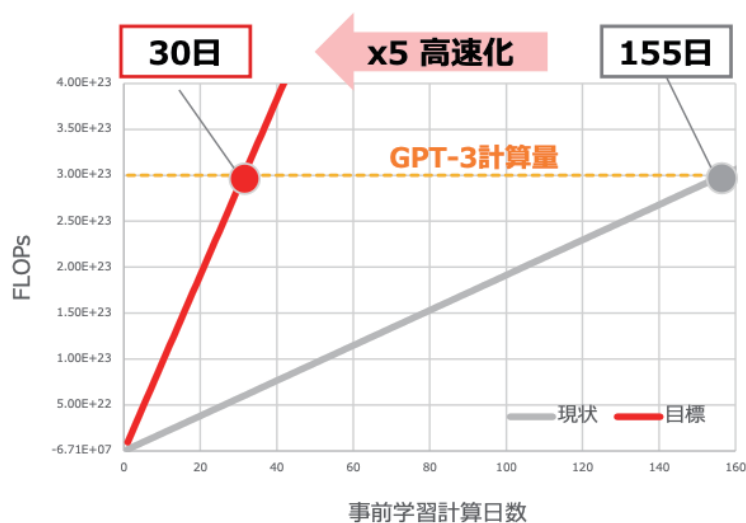
- GPT-3: 3×10^{23} FLOPs
- 従来は大規模言語モデルなど巨大モデルの学習は「富岳」の想定アプリケーションでなかったため、それに向けた最適化は施されていなかった

○「富岳」高速化

- 現状性能: 22 PetaFLOP/s
- 目標性能: **110 PetaFLOP/s**
 - 155日かかる学習を**30日**で完了

○高速化施策

- 大規模深層学習フレームワークの「富岳」への性能最適化
 - 演算の効率: 10%⇒50%の向上
(**5倍高速化**が本件の成果になる)



「富岳」全体の約1/5 (32,768ノード) を使用する
場合の試算

23

「富岳」でGPTを学習するには？

○高性能計算分野の課題

- 深層学習フレームワークMegatron-DeepSpeedの「富岳」への移植 (完了)
- 小規模な行列の積をバッチ処理する部分の高速化
- TofuDネットワーク上での集団通信の高速化
- FP16を用いても安定に学習できる方法の開発

○自然言語処理分野の課題

- 言語データの収集・クリーニング
- 弁護士による法務確認
 - 研究成果 (ソースコード・モデル・データ) の公開における著作権や契約等の制限を確認し、合法的な公開体制を構築
- 事後学習・微調整の手法の検討

24

○課題名

- 「富岳」を活用した大規模言語モデル分散並列学習手法の開発
(課題番号：hp230254)

○実施期間

- 2023年5月24日から2024年3月31日まで（「富岳」政策対応枠の利用期間）

○各機関・企業の役割

- 東京工業大学：全体総括、大規模言語モデルの並列化および高速化
- 東北大学：学習用データの収集、モデルの選択
- 富士通：大規模言語モデルの高速化
- 理化学研究所：大規模言語モデルの分散並列化・通信高速化、大規模言語モデルの高速化
- (以下8月1日に追加)
- サイバーエージェント：日本語データの提供
- 名古屋大学：分散深層学習環境の構築
- Kotoba Technologies Inc.：構築したモデルの評価

実施体制

Collaborators

GPT-Fugaku Team

GPT FUGAKU
EST. 2021

Rio Yokota

DL4Fugaku Team @ R-CCS



R-CCS



Kenichi Kobayashi

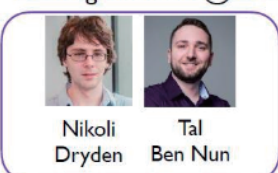


Naoto Fukumoto



Akihiko Kasagi

DL4Fugaku Team @ LLNL



Fujitsu



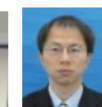
Masafumi Yamazaki



Hiroki Tokura



Takumi Honda



Tsuguchika Tabaru

Transformerの性能を「富岳」上で最適化するため、ソフトウェアスタックの各レイヤーの性能分析と最適化

特に、**密行列積の高速化**と、**通信性能の最適化**

Transformer (GPT-x)

Transformerの性能測定、ボトルネックの解析

並列化 (Megatron-DeepSpeed)

3種類の並列化を組み合わせた「富岳」向け通信性能最適化

深層学習フレームワーク (PyTorch)

富士通が高速化した「富岳」向けフレームワークを使用。LLM向け高速化実施中

数学ライブラリ

密行列積ライブラリのTransformer向け高速化

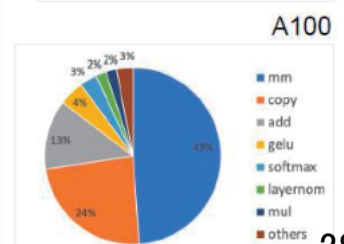
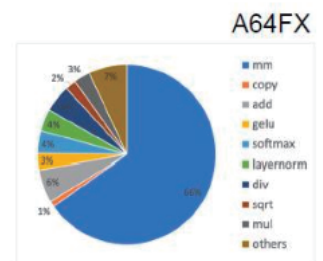
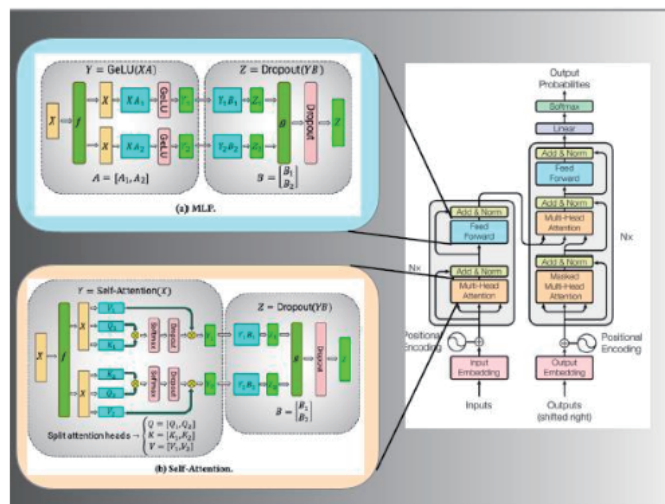
TransformerのA64FX向け性能最適化(演算部分)

FLOPsの99%は小規模な密行列積

→ A64FXでは66%、A100では49%の時間がこの部分に費やされている

→ 現時点で理論ピークの1/3の性能になっており大幅に向上できる可能性がある

ある共通の入力 X に対しそれぞれの変換行列を適用して $Q=XW_Q, K=XW_K, V=XW_V$ を用意する。自分自身の要素との注目度合いを抽出する。



データ並列

データは分散
モデルは冗長
勾配の集約通信

課題：バッチサイズの増大に伴う汎化性能の低下

解決策：正則化・最適化手法の工夫

テンソル並列

データは冗長
モデルは分散
活性の集約通信

課題：通信頻度の増加

解決策：通信のオーバーラップ

パイプライン並列

データは冗長
モデルは分散
活性の1対1通信

課題：パイプラインバブル

解決策：パイプラインの工夫

分散並列化性能

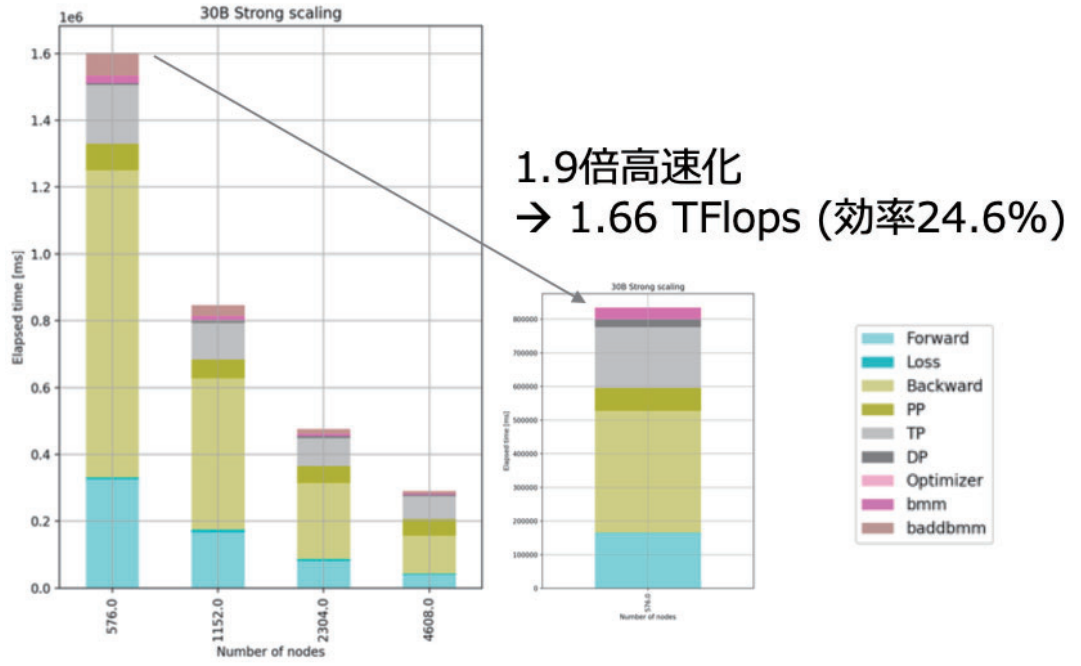
pytorch1.13(default)
parameters=30B

pipeline parallelism=48
tensor parallelism=6
data parallelism=[2,4,8]

num-layers 48
hidden_size 7168
num-attention-heads 56

micro-batch-size 1
global-batch-size 1536

Initial FLOP/s/node 0.7
↓ x2
Current FLOP/s/node 2.3



LLM開発に対する所感



大規模であればあるほど良い (Scaling Law)

→ パラメータ数だけでなく、学習データも巨大でなければならない

大きいものを一つ作ったところで1年で陳腐化する → 学習したモデル自体が成果物ではない

→ ノウハウの蓄積、人材の育成が重要

今は各社がバラバラにLLMを学習しているが、今後統合されていくことが望ましい

→ 計算資源 ⇔ データ ⇔ 人材 のバーター

→ ごく一部を除いては、fine-tuning、prompting、APIを使った独自応用 をやるべき

国外で開発されているオープンなモデルを活用しないのはもったいない

→ LLaMA2などから日本語と英語でさらに継続学習

31



Tokyo Tech

Thank You



3次元形状生成AIの産業応用に向けて

名古屋大学 大学院工学研究科 土木工学専攻 構造・材料工学講座 准教授
理化学研究所 計算科学研究センター 複雑現象統一的解法研究チーム 客員研究員

西口 浩司

kojinishiguchi@civil.nagoya-u.ac.jp

1/74

西口 浩司 (にしぐち こうじ)

• 専門分野：**計算力学（非線形構造解析，構造-流体連成問題，高性能計算）**

• 経歴：

—1985年：広島県広島市生まれ

—2004年～2010年：広島大学 工学部（建設・環境系）と修士課程

—2011年～2015年：広島大学 博士課程

—2010年～2016年：日東電工株式会社 研究開発本部

✓ CAEに関わる研究開発（4年），新規事業立案（2年）

—2016年～2020年：理化学研究所 計算科学研究センター

✓ 特別研究員，スパコン「京」PJ，スパコン「富岳」PJの研究

—2020年～2023年：名古屋大学 大学院工学研究科 土木工学専攻 講師

—2023年～現在：名古屋大学 大学院工学研究科 土木工学専攻 准教授

✓ 客員：理化学研究所，神戸大学，山梨大学



2/74



もくじ

NAGOYA UNIVERSITY

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

3/74



3次元形状生成AIの研究を始めた動機

NAGOYA UNIVERSITY

巨大アルミダイキャストによる自動車構造の革新



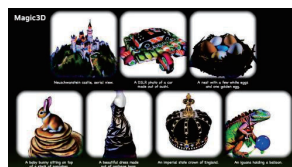
質量：30%減，製造コスト：40%減



ギガプレス（テスラ）



**深層生成モデル（Parameter-to-3D）によって
ギガキャスト構造の最適設計は可能か？**



Magic3D（NVIDIA，2022年11月）



Shap-E（OpenAI，2023年5月）

3次元形状生成AIやTransformerに基づくLLMの急速な発展

4/74



NAGOYA UNIVERSITY

主要な研究メンバー



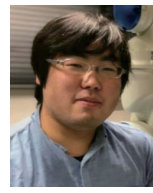
 名古屋大学

准教授 西口 浩司



 名古屋大学

M1 戸井田 一聖 君



 東北大学
TOHOKU UNIVERSITY

助教 千葉 直也 先生



 東京工業大学
Tokyo Institute of Technology

助教 和田 有司 先生

5/74



NAGOYA UNIVERSITY

もくじ

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

6/74

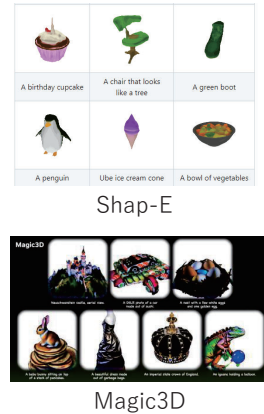


3D生成AI研究の活発化

NAGOYA UNIVERSITY

- StableDiffusionのような画像生成AIだけでなく、3次元形状を生成するAIも2022年以降、続々と登場。
- ShapeNet等の3次元形状とテキストのデータセットを、深層生成モデルに学習させたもの。
 - 自然言語・画像・動画に比べて、**3次元形状のデータセット数は圧倒的に少ない。**
 - **構造力学的設計に応用可能なデータセットは未だ提案されていない。**

モデル名	発表年月	研究グループ	3D形状表現	モデル・アーキテクチャ	データセット	3Dデータ数
Shap-E	2023年5月	OpenAI	陰関数	Transformerベースの拡散モデル	ShapeNet (3D形状), WebImageText (2D画像)	数百万個
Point-E	2022年12月	OpenAI	3次元点群	Transformerベースの拡散モデル	ShapeNet (3D形状), WebImageText (2D画像)	数百万個
Magic3D	2022年11月	NVIDIA	3Dメッシュ	NeRF, 拡散モデル	COCO (2D画像), ImageNet (2D画像)	なし
DreamFusion	2022年9月	Google, UC Berkeley	陰関数	NeRF, 拡散モデル	COCO (2D画像), ImageNet (2D画像)	なし



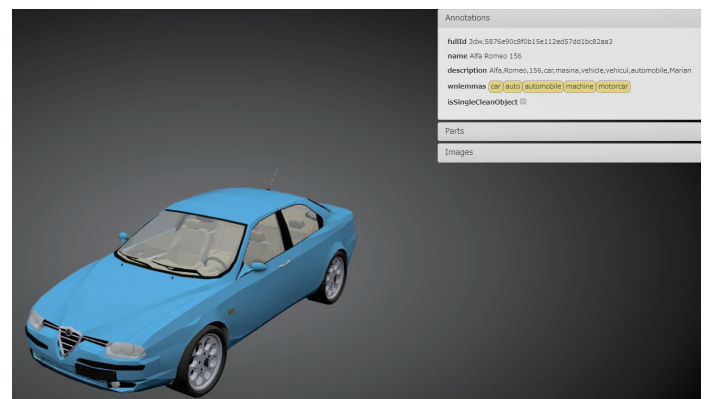
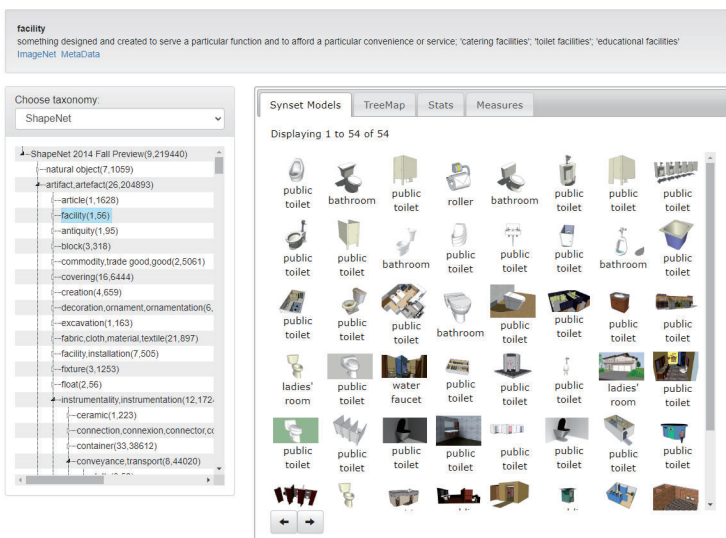
7/74



ShapeNet

NAGOYA UNIVERSITY

- 3次元形状の大規模データセット（現時点で約922万個）
- コンピュータグラフィックスやコンピュータビジョンに関連する研究が主な用途
- 構造力学的設計を想定したデータセットではない

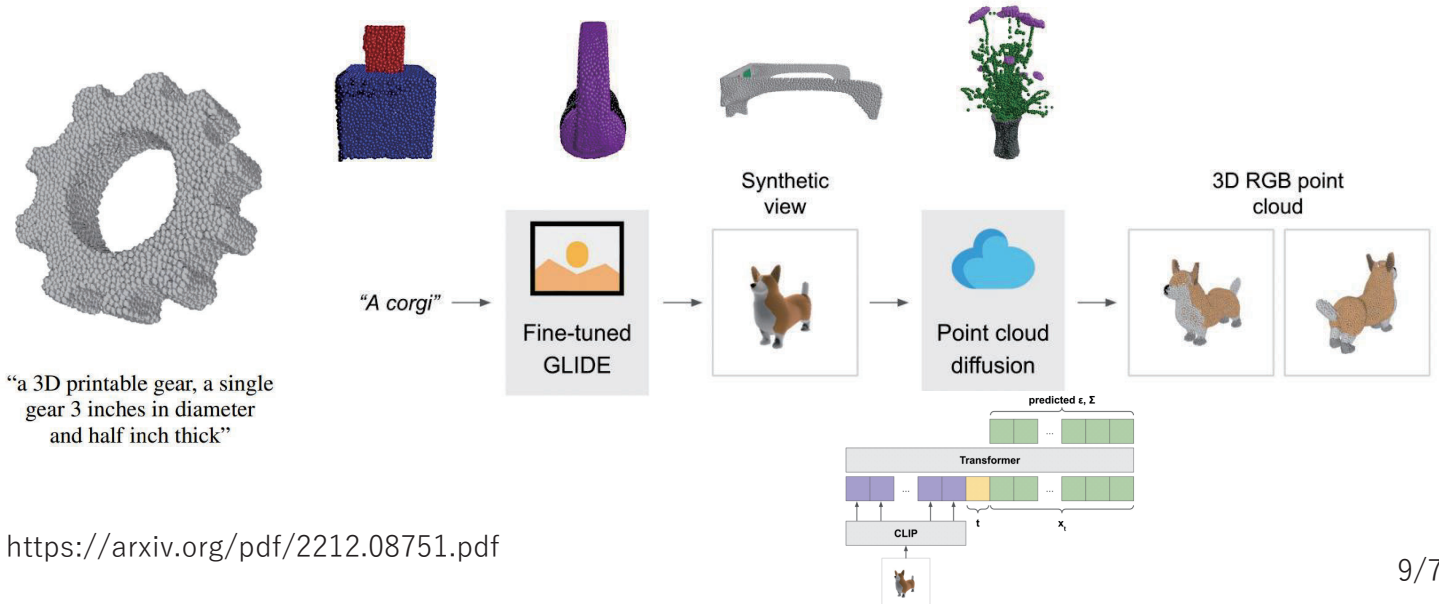


<https://shapenet.org/taxonomy-viewer>

8/74

Point-E (2022年12月, OpenAI)

- 2次元画像生成モデル(text-to-image)も使用されている。
- 形状パラメータから3次元形状を生成できる可能性が示されている。



もくじ

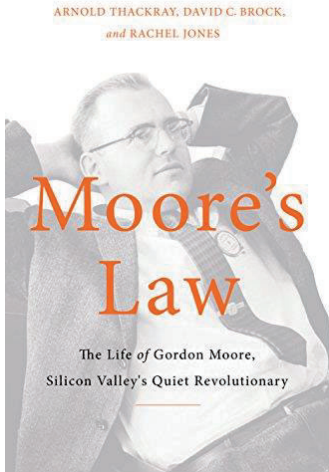
1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望



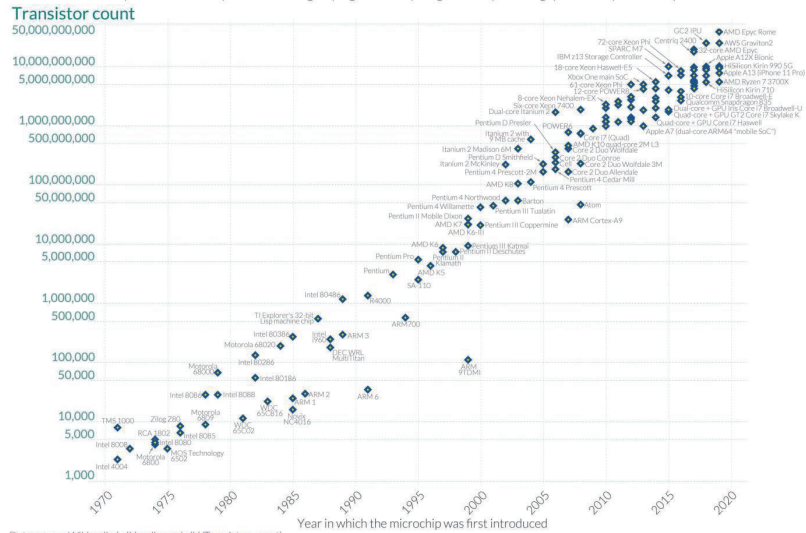
ムーアの法則 (2年で2倍)

NAGOYA UNIVERSITY

- コンピューターチップ上のトランジスタ数のスケール則：2年で2倍
- 1965年にインテルの共同創設者ゴードン E. ムーアによって提唱



Moore's Law: The number of transistors on microchips has doubled every two years
 Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.



11/74

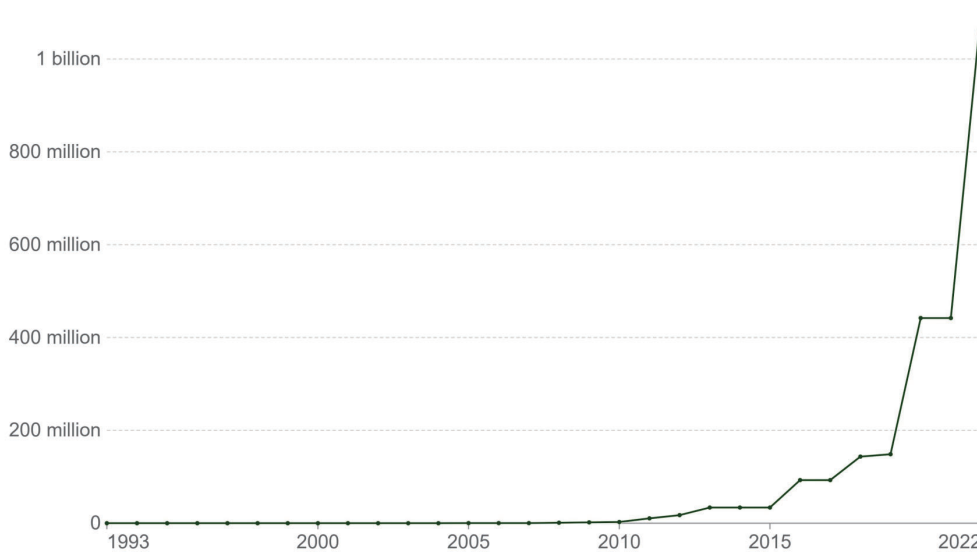


世界最速スパコンの計算性能の推移

NAGOYA UNIVERSITY

Computational capacity of the fastest supercomputers

The number of floating-point operations¹ carried out per second by the fastest supercomputer in any given year. This is expressed in gigaFLOPS, equivalent to 10⁹ floating-point operations per second.



Source: TOP500 Supercomputer Database (2023)

OurWorldInData.org/technological-change • CC BY

1. Floating-point operation: A floating-point operation (FLOP) is a type of computer operation. One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.

12/74



2020年のスマホ ≡ 2001年の世界最速スパコン

NAGOYA UNIVERSITY

- 指数的变化は直感的に理解しにくい。
- 技術史の観点では、指数関数的な性能向上は異質。
- コンピュータ性能の指数関数的向上がなければ、最近の生成AIの急速な発展はありえない。

iPhone 12



スーパーコンピュータ「ASCI White」 (米国)



https://en.wikipedia.org/wiki/ASCI_White

理論性能値：約11TFlop/s*

*<https://blog.adobe.com/en/publish/2022/11/08/fast-forward-comparing-1980s-supercomputer-to-modern-smartphone>

理論性能値：約12TFlop/s**

**<https://www.top500.org/lists/top500/2001/11/>

13/74

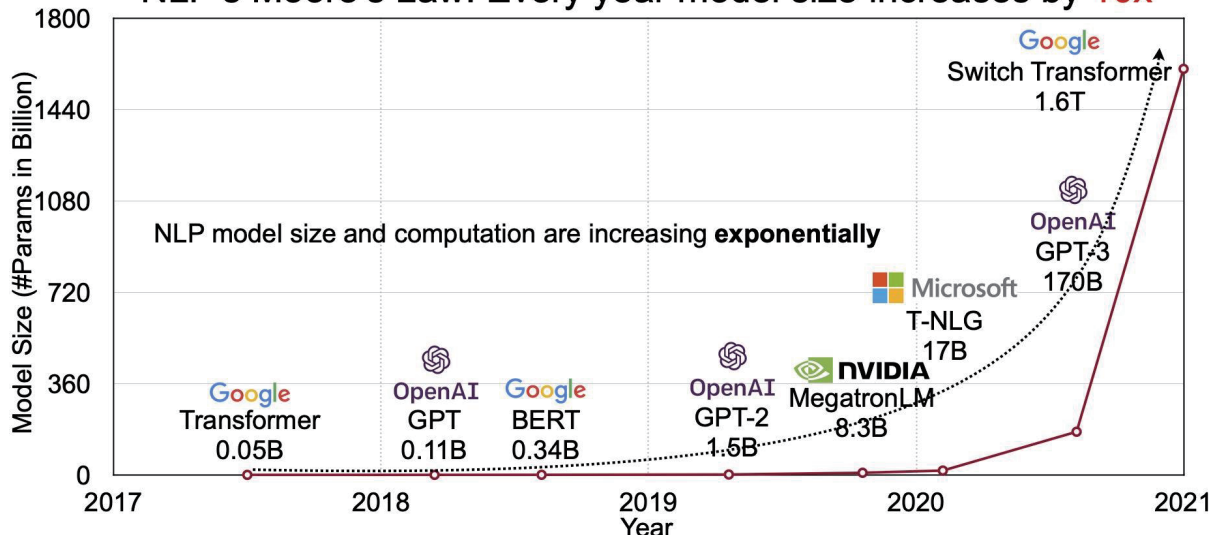


LLMのスケーリング則 (1年で10倍)

NAGOYA UNIVERSITY

- 超並列計算に適したディープニューラルネットワーク，Transformerの登場
- 1年で10倍のパラメータ数の向上：CAEにも革新を起こせないだろうか？

NLP's Moore's Law: Every year model size increases by 10x



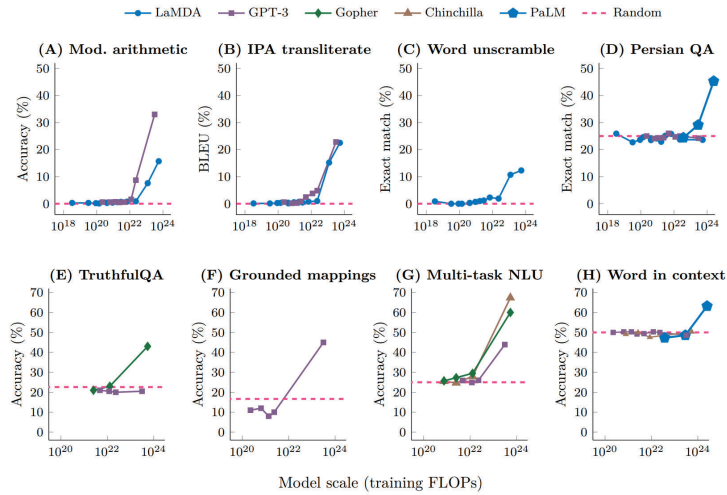
https://hanlab.mit.edu/projects/efficientnlp_old/assets/nlp_trend.jpg

14/74



Emergent Abilities at 10^{22} - 10^{24} FLOPs

- Scaling up large language models (LLMs) has been shown to predictably improve performance at 10^{22} - 10^{24} FLOPs
- 学術的興味：3次元形状生成モデルでも創発性は観測されるか？



Emergent Abilities of Large Language Models, <https://arxiv.org/abs/2206.07682>



VUCA時代のR&D戦略に必要なこと

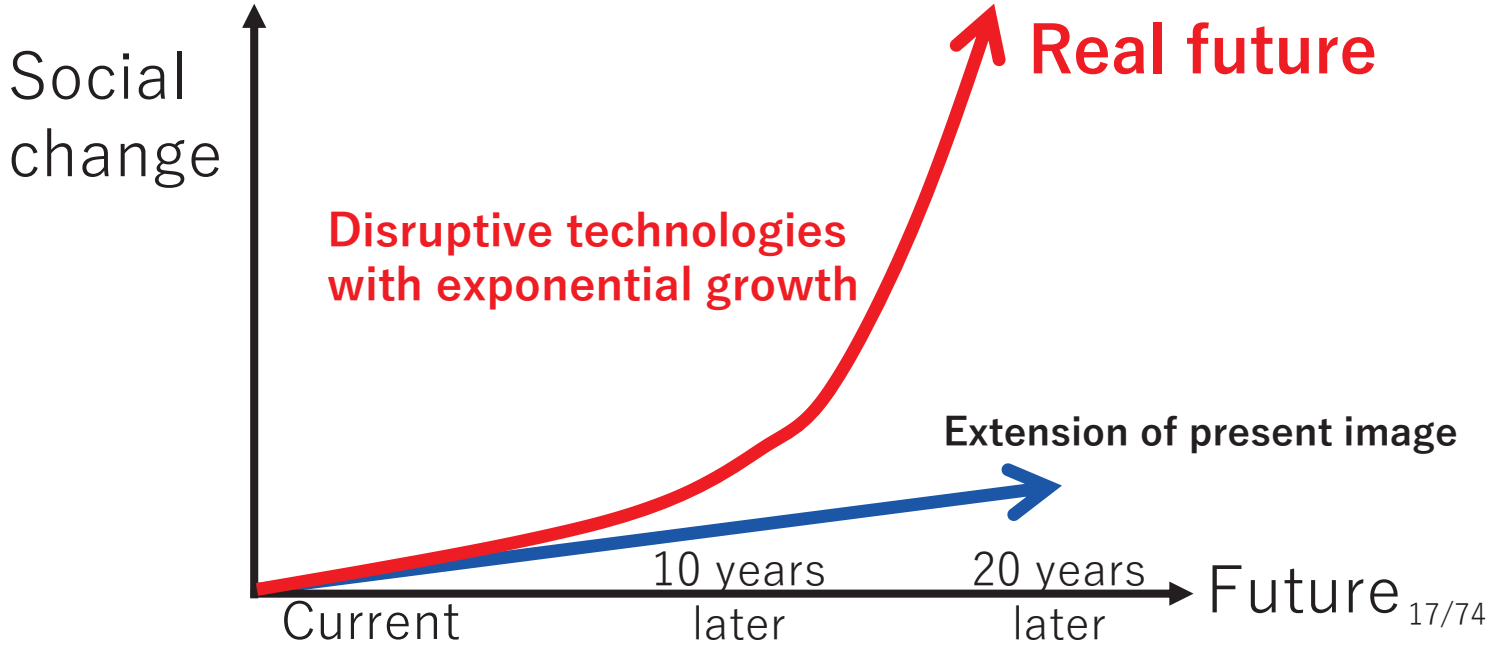
- 現代は、VUCA (Volatility・Uncertainty・Complexity・Ambiguity) 時代と言われる
 - 社会やビジネスにおいて未来の予測が難しくなる状況
 - しかし、高い確度で予測できることはある
- PEST分析
 - 外部環境を政治 (Politics) ・経済 (Economy) ・社会 (Society) ・技術 (Technology) の4つの要因に分類し、自社に与える影響を読み解く分析手法
 - 高い確度で予測できるもの
 - ✓ 社会 (Society) : 人口動態
 - ✓ 技術 (Technology) : ムーアの法則 (2年で2倍) , LLMスケーリング則 (1年で10倍)
- 企業のR&D戦略において「2つのスケーリング則」を考慮することは必要不可欠
 - ✓ 高い確度で予測できるものを軸にR&D戦略を立てれば、流行りのキーワード (DX, 生成AI, 量子コンピュータ等) に振り回されることも少なくなる。
 - ✓ スパコン「富岳」水準のコンピュータがコモディティ化する未来を想定されていますか？



VUCA時代のR&D戦略に必要なこと

NAGOYA UNIVERSITY

- 高度にデジタル化された社会では，社会も指数関数的に変化する。
- 本当の未来は，現状の線形的な延長線上には存在しない。



もくじ

NAGOYA UNIVERSITY

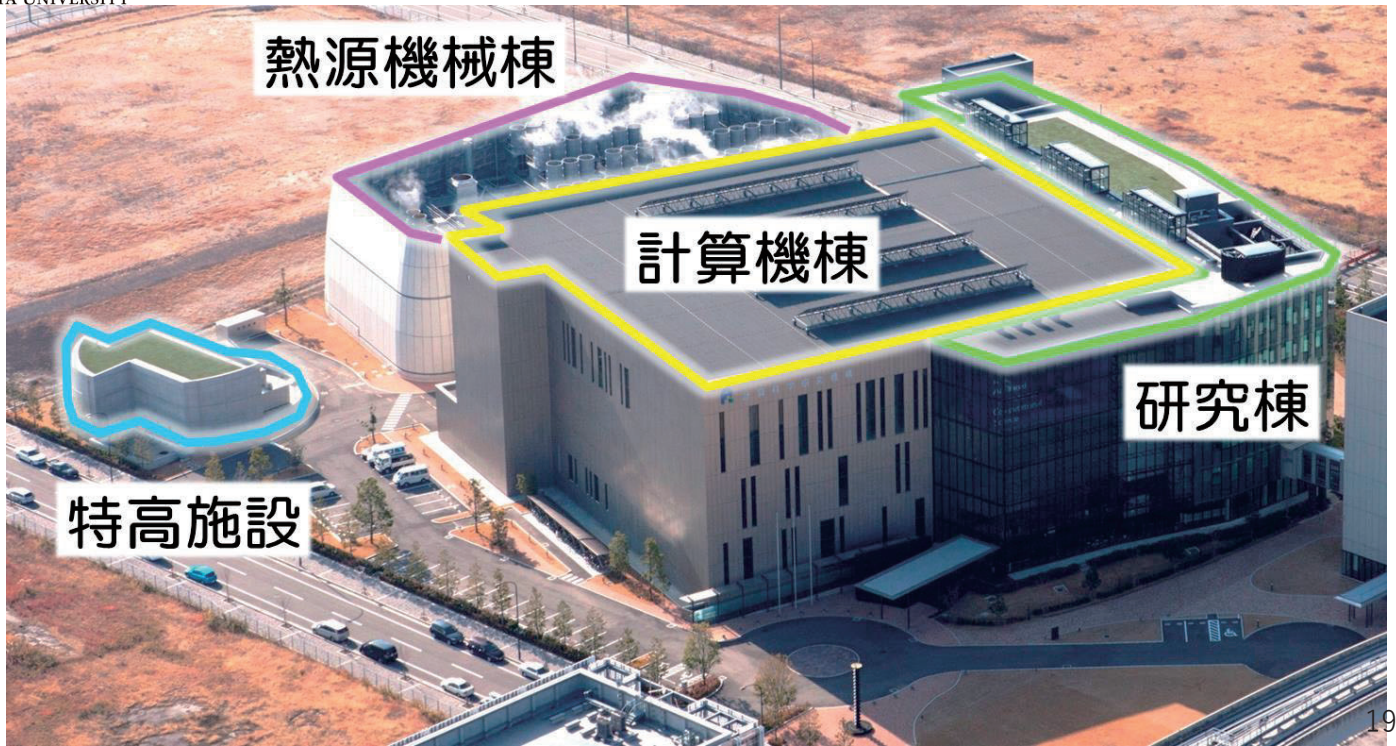
1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

18/74



NAGOYA UNIVERSITY

理化学研究所 計算科学研究センター



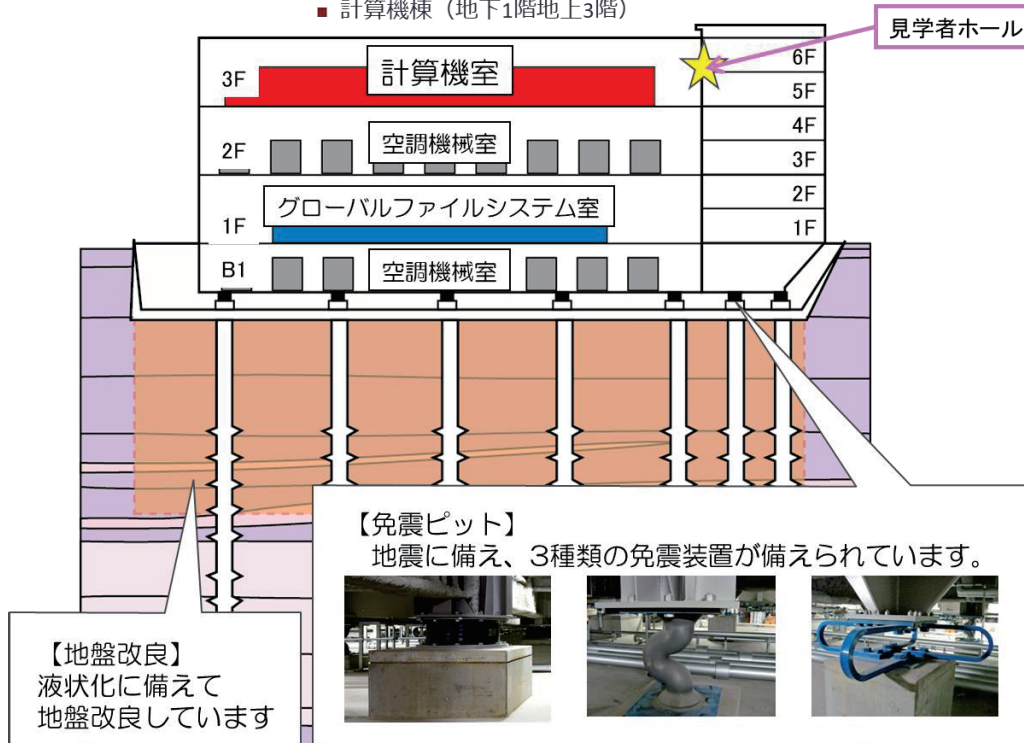
19/74



NAGOYA UNIVERSITY

理化学研究所 計算科学研究センター

■ 計算機棟 (地下1階地上3階)



20/74



スーパーコンピュータ「富岳」

NAGOYA UNIVERSITY



21/74



スーパーコンピュータ「富岳」の特長

NAGOYA UNIVERSITY

歴代1位と1秒間の計算回数



シミュレーション

AI

ビッグデータ

- CPU (A64FX) マシン
- GPUマシンでは難しい, 高い汎用性を実現
- 各種ベンチマークで4冠 (2020年時点) を獲得したのは史上初

出典：2020年6月23日日経新聞朝刊第一面

22/74



そもそも、スパコンとは？

NAGOYA UNIVERSITY

- 明確な定義はない.
- 現在, 最高レベルの計算性能を持つコンピュータを指す.
- 現在のスパコン≒超並列コンピュータ
 - ▶ たくさんのコンピュータを通信ケーブルでつないでいる



23/74



なぜスパコンは計算が速いのか？

NAGOYA UNIVERSITY

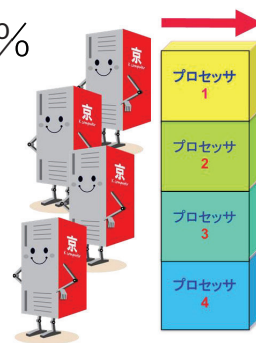
- 1 台のコンピュータで要する計算時間は, 4 台のコンピュータを使うと, 理想的には 4 分の 1 になる

= 並列化効率100%という状態



- しかし, 実際には並列化効率は100%にならない.

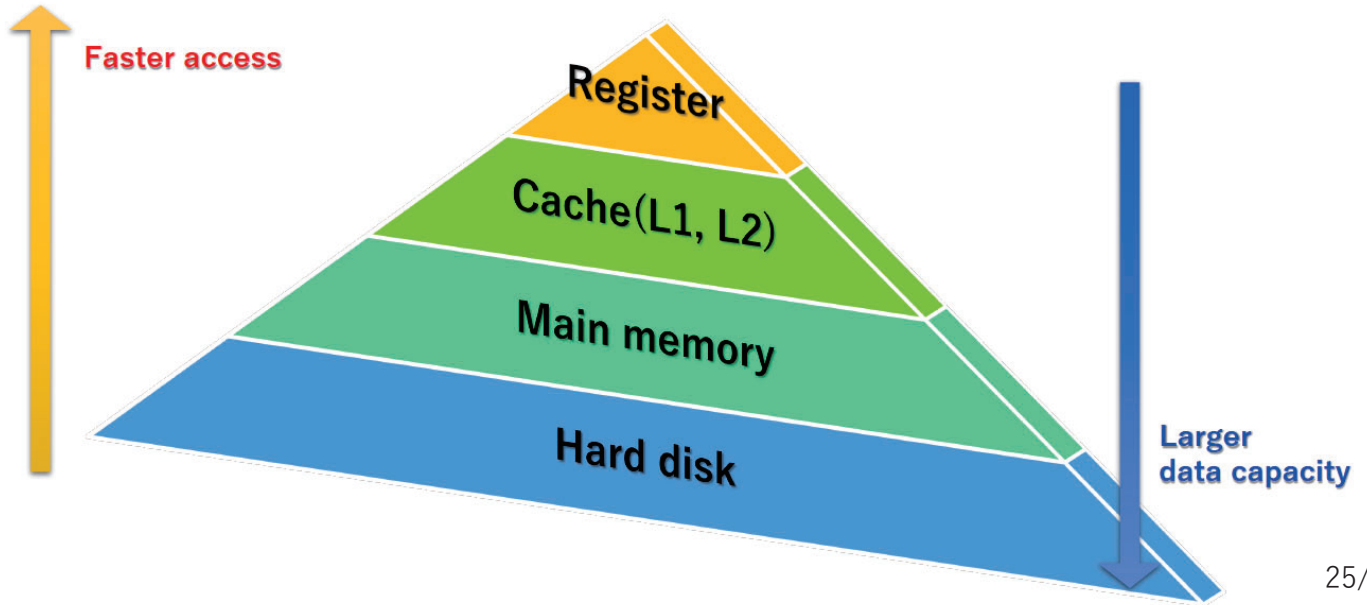
- 並列化効率を上げるには
 - ▶ 均一に計算を分担させる
 - ▶ メモリアクセスを速くする
 などの工夫が必要.



24/74

コンピュータのメモリ階層

- アクセスが高速な領域をできるだけ無駄なく使えるように、アルゴリズムを設計することが必要。



25/74

「富岳」と第二階層スパコン

- 革新的ハイパフォーマンス・コンピューティング・インフラ(HPCI)：国内の大学や研究機関の計算機システムやストレージを高速ネットワークで結んだ共用計算環境基盤。
- フラッグシップマシン「富岳」の第二階層として複数のスーパーコンピュータがHPCIで利用可能
- RISTの奥田基さんのご説明：
 - 「産業利用は、京のときから始まったものであるが、学術利用に留まらず、産業利用も積極的に推進する制度は、現時点でも、世界にはほとんどない。日本のHPCIと同様の環境として、米国ではINCITEやXSEDE、欧州ではPRACEやEuro HPCがあるが、HPCIは世界に類を見ないスーパーコンピューティング環境である」
 - <https://ascii.jp/elem/000/004/078/4078861/2/>

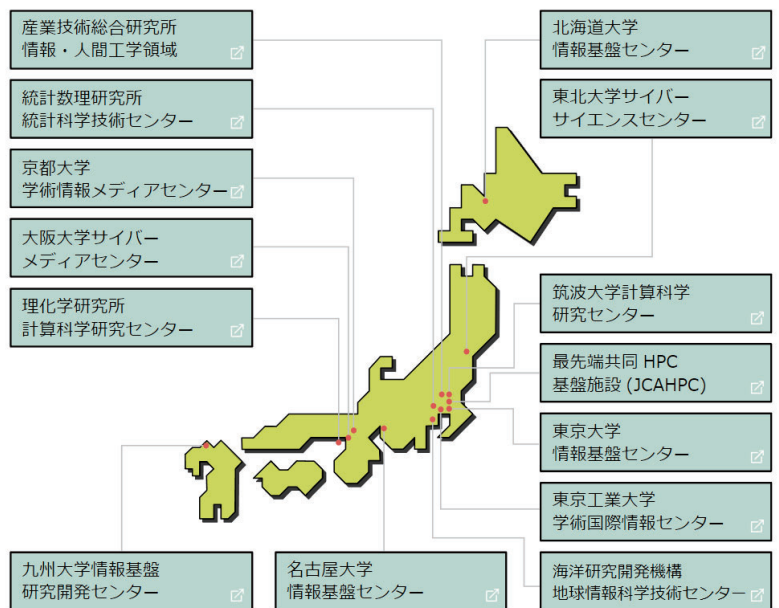


図1：HPCIシステムを構成する計算機資源を提供するシステム構成機関 (2023年度)

https://www.hpci-office.jp/about_hpci/what_is_hpci

26/74



もくじ

NAGOYA UNIVERSITY

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. **ギガキャスト構造の利点と設計上の課題**
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

27/74



ギガキャスト構造のインパクト

NAGOYA UNIVERSITY

- SpaceX社の冶金技術（アルミニウム）を活用したTeslaのギガキャストの登場
 - 2020年、Teslaは主力BEV「モデルY」の後部の車体構造で初採用
 - **質量：30%減、製造コスト：40%減**
 - 幾何学的自由度が高い構造設計が可能
- 2022年、NIOやXPeng等の中国を代表するBEVメーカーに加えてボルボ社も採用を発表。
- 2023年6月、トヨタ自動車も採用を発表。



テスラのギガキャスト構造
(アルミニウム)

<https://thelastdriverlicenseholder.com/2022/04/20/tesla-revenue-grows-81-percent-in-q1-2022/>



テスラのギガキャスト装置
(アルミニウムのダイキャスト装置)

<https://lowcarb.style/2022/08/23/tesla-gigapress-giga-texas/>

28/74

• <https://jp.reuters.com/economy/industry/FMOAA43NZRQVREP7A2LDHDPGE-2023-09-19/>

• 引用：

- ▶ テスラはプラットフォーム（車台）と呼ばれる**複雑な車体下部のほぼ全てを一体成型**する革新的な技術の実用化に近づいている。自動車の下部は約400点の部品で構成されており、これが1つになればテスラは競合他社とのリードをさらに広げることになる。
- ▶ 米ケアソフト・グローバルのテリー・ウォイクウスキー社長は、もしテスラが車体下部のほとんどをギガキャスト化することに成功すれば、自動車の設計・製造方法の変革がさらにと進むとみる。「ものすごく強力な手法だ。業界に大きな影響を及ぼし得るが、非常に困難な課題もある」と語る。「 castingはかなり難しく、特に大型で複雑なものほど困難だ」と言う。
- ▶ 関係者5人によると、テスラが2020年代半ばまでに2万5000ドルで発売しようとしている小型EVには、前部と後部、そしてバッテリーが搭載される中央の底部を組み合わせた単一の大きなフレームが使われる可能性がある。

企業・産業 焦点：米テスラ「ギガキャスト」高度化へ、静かに進む技術革新

Norihiko Shirouzu

2023年9月19日 午後 6:29 GMT+9・2日前更新



9月14日、米電気自動車（EV）大手テスラは、車体の主要構造部分を1回のダイカスト鋳造プレスで成型する「ギガキャスト（GigaCasting）」と呼ばれる技術のパイオニアだ。写真は同社のイーロン・マスクCEO。パリで6月撮影（2023年 ロイター/Gonzalo Fuentes）

【オースティン（米テキサス州） 14日 ロイター】 - 米電気自動車（EV）大手テスラ（TSLA）は、車体の主要構造部分を1回のダイカスト鋳造プレスで成型する「ギガキャスト（GigaCasting）」と呼ばれる技術のパイオニアだ。この手法は生産の効率化や生産コストの削減につながることからライバル企業はテスラに追いつこうと必死だが、テスラがさらなる高度化を進めていることが関係者5人の話で明らかになった。

29/74

ギガキャスト構造設計の課題

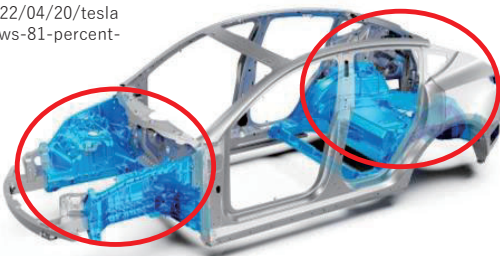
① 巨視的スケール：クラッシュャブル・ゾーンの最適構造の探索

- トポロジー最適化
 - ✓ 線形問題（弾性体の微小変形問題）はほぼ確立されている。
 - ✓ 大変形・破断・弾塑性を考慮できる手法は確立されていない。

② 微視的スケール：ランダムに発生する鑄巣の数理モデル化

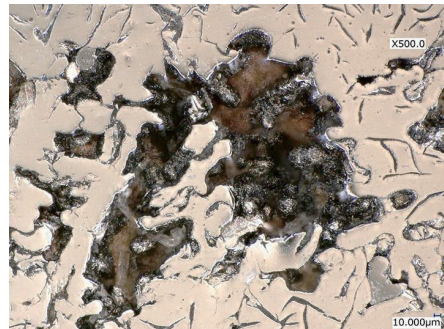
- 応力ひずみ曲線のばらつき
- 破断ひずみのばらつき

<https://thelastdriverlicenseholder.com/2022/04/20/tesla-revenue-grows-81-percent-in-q1-2022/>

**クラッシュャブル・ゾーン**

(衝突のエネルギーを吸収し、乗員を保護)

<https://www.keyence.co.jp/ss/products/microscope/vhx-casestudy/automobile/die-casting.jsp>

**鑄巣**

(溶融金属内の空気やガスの混入等による空洞) 30/74



もくじ

NAGOYA UNIVERSITY

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. **3次元形状のデータセット生成方法**
 - **階層型の固定直交メッシュ法**
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

31/74



データセット生成の方針

NAGOYA UNIVERSITY

- 3次元形状と力学的パラメータ（寸法，質量，剛性，ひずみエネルギー等）がペアになったデータセットを作る。
- **3次元形状をどのように作るのか？**
 - ▶線形トポロジー最適化において，境界条件境界条件・制約条件・目的関数に変動を与えることで，多ケースの3次元形状と線形領域の力学的パラメータ（質量・剛性等）を求める。
- **非線形領域の力学的パラメータはどのように求めるか？**
 - ▶オイラー型構造-流体統一解法で，線形トポロジー最適化で求めた3次元形状の弾塑性解析を行うことで，非線形領域の力学的パラメータ（強度・大変形時の弾塑性ひずみエネルギー等）を求める。

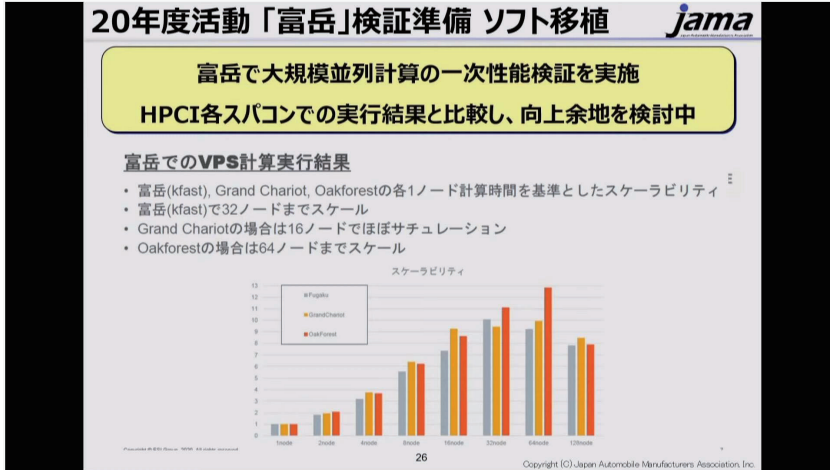
32/74



従来のな弾塑性衝突解析の課題①：並列化効率

NAGOYA UNIVERSITY

- 商用の動的陽解法FEMコードによる弾塑性衝突解析では、「富岳」の32ノードで並列化効率が限界に達した[1]（富岳は158,976ノードを有する）。
- スパコン「京」用にチューニングされたLS-DYNAでも、768ノードで並列化効率が限界に達した[2]（京は88,192ノードを有する）。接触解析を行えば並列化効率はさらに悪化する[2]。



[2] Kondo K, Minami K, Hasegawa Y, et al. Performance evaluation using LS-DYNA hybrid version on the K Computer. In: Proceedings of 9th European LS-DYNA Conference; 2013; Manchester, UK.

共用開始記念イベント HPCIフォーラム～スーパーコンピュータ「富岳」への期待～

[1] HPCIを活用した自動車先端シミュレーション技術検証, スーパーコンピュータ「富岳」共用開始イベントHPCIフォーラム～スーパーコンピュータ「富岳」への期待～

33/74



従来のな弾塑性衝突解析の課題②：計算メッシュ生成

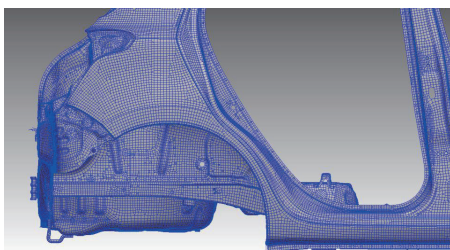
NAGOYA UNIVERSITY

- 計算メッシュ生成に多大な時間と労力を要する（シェル構造に基づく従来の車体構造）
- ギガキャスト構造の計算メッシュ生成は、シェル構造よりもさらに難しい
 - 巨視的スケール：幾何学的に複雑な形状に対する計算メッシュ生成
 - 微視的スケール：構造内部の鑄巣形状に対する計算メッシュ生成

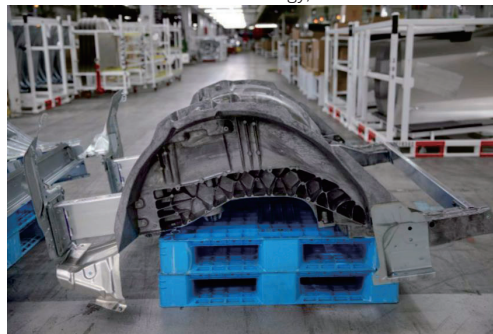
西口浩司, et al. "ラグランジュマーカー粒子を用いたオイラー型有限体積法による圧縮性固体解析." 土木学会論文集 A2 (応用力学) 75.2 (2019): 1_237-1_248.

<https://blog.evsmart.net/electric-vehicles/tesla-aluminum-die-cast-technology/>

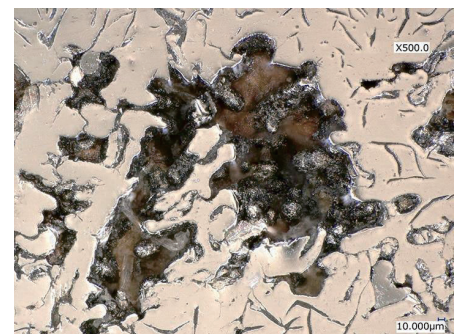
<https://www.keyence.co.jp/ss/products/microscope/vhx-casestudy/automobile/die-casting.jsp>



シェル構造の計算メッシュ



テスラ/モデルYのギガキャスト構造



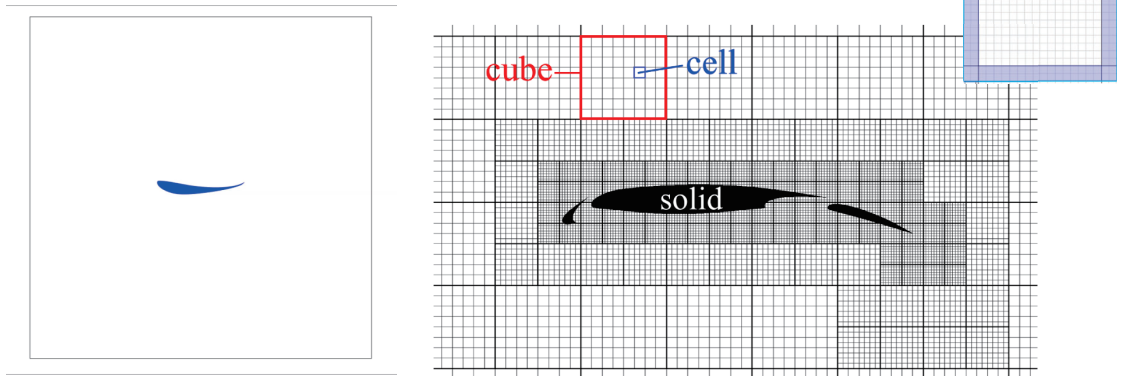
鑄巣

34/74



Building-cube method

- Computational domain is divided into cubic regions (cube), which have various sizes to effectively capture complex geometries.
- Each “cube” has same number of cells, and loop processing is conducted within each “cube”. Thus it is easy balance computational load.
- Data exchange is required between adjacent cubes.



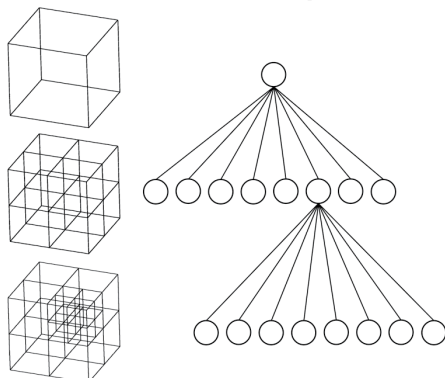
* Nishiguchi K et.al, M. Full Eulerian deformable solid - fluid interaction scheme based on building-cube method for large - scale parallel computing. *Int J Numer Methods Eng.* 2019; 117: 221– 248. <https://doi.org/10.1002/nme.5954>



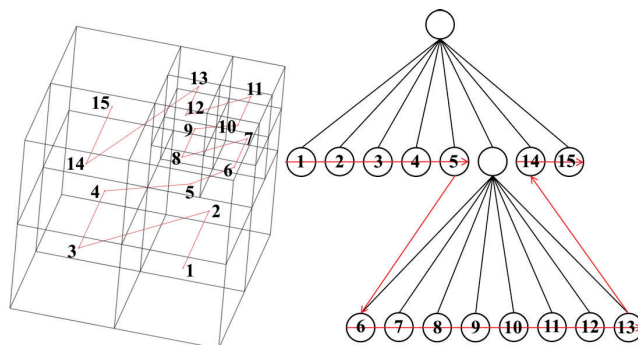
Data structure of “cube”

- Data structure
 - **Octree**
- Spatial index of the cube
 - stored in a one-dimensional array using **Z-ordering** (space-filling curve) to simplify the loop processing algorithm.

Octree concept



Spatial index of the cube



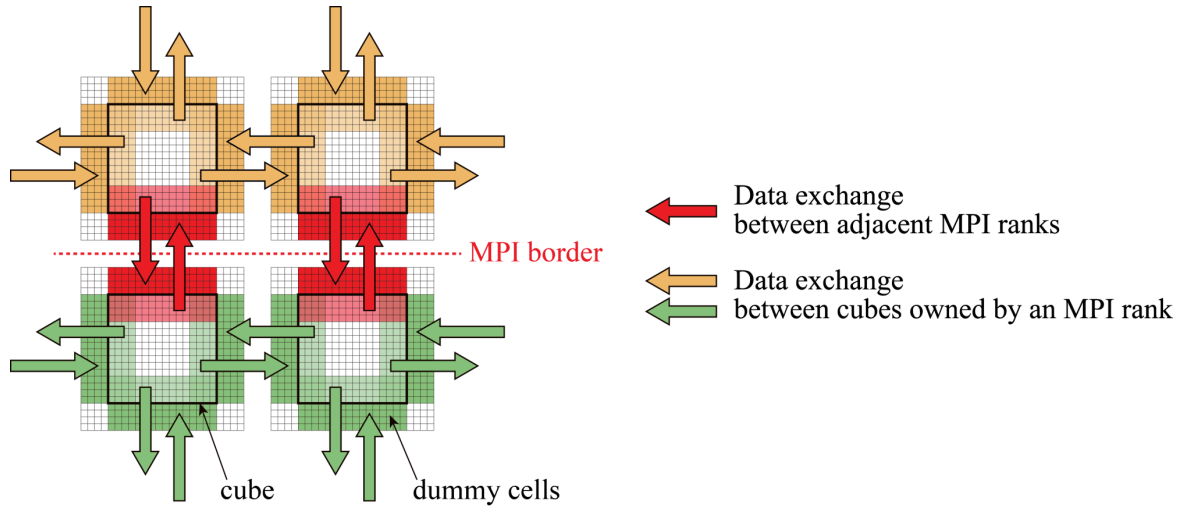
* Nishiguchi K et.al, M. Full Eulerian deformable solid - fluid interaction scheme based on building - cube method for large - scale parallel computing. *Int J Numer Methods Eng.* 2019; 117: 221– 248. <https://doi.org/10.1002/nme.5954>



Data exchange between "cubes"

NAGOYA UNIVERSITY

- In BCM, data is exchanged not only globally between adjacent MPI ranks but also locally between cubes owned by an MPI rank.
- In the present study, the width of the dummy cells is four cells.



* Nishiguchi K et.al, M. Full Eulerian deformable solid - fluid interaction scheme based on building - cube method for large - scale parallel computing. *Int J Numer Methods Eng.* 2019; 117: 221–248. <https://doi.org/10.1002/nme.5954>

37/74

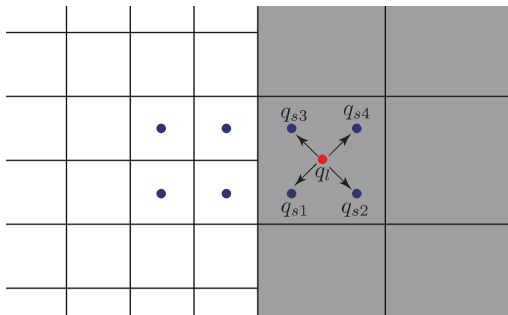


Interpolation for dummy cells

NAGOYA UNIVERSITY

- Interpolation is used to exchange data in dummy cells (halo region) between Cubes of different sizes.
- Exchange data: velocity, pressure, etc.

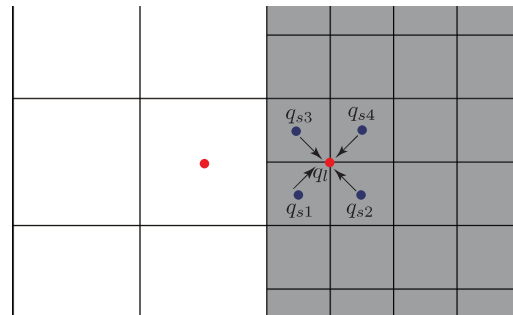
Large cube to small cube



$$q_{s1} = q_{s2} = q_{s3} = q_{s4} = q_l$$

zero-order interpolation

Small cube to large cube



$$q_l = \frac{q_{s1} + q_{s2} + q_{s3} + q_{s4}}{4}$$

first-order interpolation

* Nishiguchi K et.al, M. Full Eulerian deformable solid - fluid interaction scheme based on building - cube method for large - scale parallel computing. *Int J Numer Methods Eng.* 2019; 117: 221–248. <https://doi.org/10.1002/nme.5954>

38/74

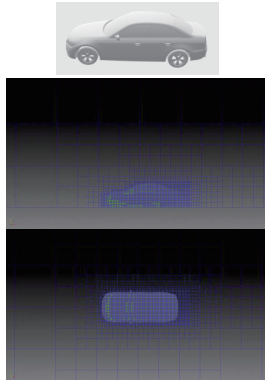


「富岳」における良好な並列化効率

NAGOYA UNIVERSITY

- スパコン「富岳」の**51200ノード（245万7600コア）**の計算で、**74.11%**の並列化効率を実現
- **682億要素**メッシュによる構造(剛体)-流体連成解析
 - 単純化された問題ではなく実問題（複雑形状を含む問題）で計測

パラメータ	値
ソルバー種別	非圧縮性
ドメインサイズ	18.75m x 8.75m x 3.75m
キューブ数	32,341
キューブあたりセル数	16 x 16 x 16 = 4,096
総セル数	32,341 x 4,096 = 132,468,736
最小セルサイズ	1.2207 mm
時間ステップ幅	2 x 10 ⁻⁶ sec
積分時間	2 x 10 ⁻⁴ sec (100 steps)
レイノルズ数	2.572 x 10 ⁶
時間横分スキーム	Crank-Nicolson
Poissonソルバー	Red-Black SOR
ノードあたりプロセス数	4
プロセスあたりスレッド数	12



ノード	キューブ数	セル数	経過時間	スケーリング
100	32,341	132,468,736	22.92 秒	100.00 %
800	260,331	1,066,315,776	32.63 秒	70.23 %
6,400	2,082,648	8,530,526,208	33.46 秒	68.49 %
51,200	16,661,184	68,244,209,664	30.927 秒	74.11 %

Ando, Kazuto, et al. "Digital transformation of droplet/aerosol infection risk assessment realized on "Fugaku" for the fight against COVID-19." *The International Journal of High Performance Computing Applications* 36.5-6 (2022): 568-586.

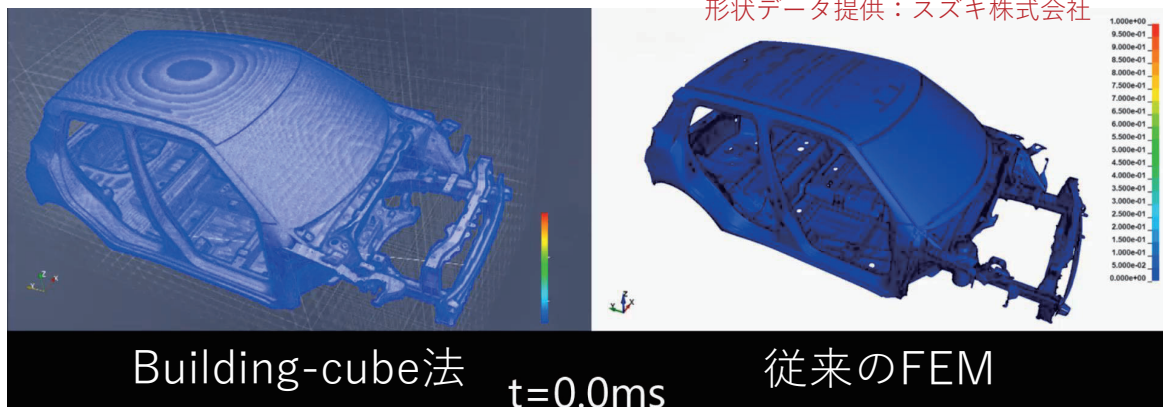
39/74



高速な計算メッシュ生成

NAGOYA UNIVERSITY

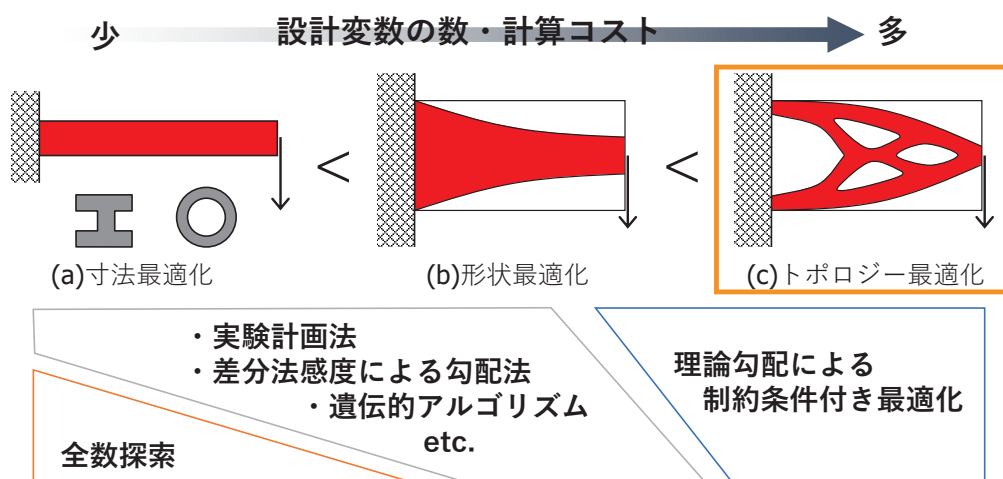
	Building-cube法（連続体要素）	従来のFEM（構造要素）
計算メッシュの作成	10分	数週間
計算実行	2.4時間	0.75時間
計算結果の見える化	1時間	1時間
合計時間	3.6時間	数週間



Nishiguchi K et al, Eulerian finite volume formulation using particle-in-cell method for large-scale parallel simulation of complex structures, *Proceedings of 4th International Conference on Multi-scale Computational Methods for Solids and Fluids*, pp.223-226, 2019.

40/74

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. **3次元形状のデータセット生成方法**
 - 階層型の固定直交メッシュ法
 - **線形範囲：トポロジー最適化**
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

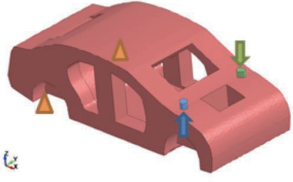


トポロジー最適化

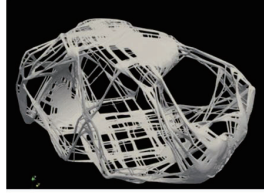
- 設計空間内の材料の有無を自由に指定し、最適な構造を模索する
- 線形材料（弾性体）の静的な微小変形問題に対する手法は、ほぼ確立されている。
- **非線形材料（弾塑性体等）の動的な大変形問題に対する手法は、確立されていない。**

Building-cube法に基づく線形トポロジー最適化

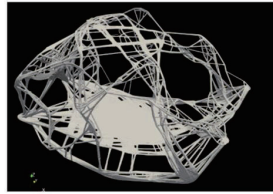
設計領域・荷重条件



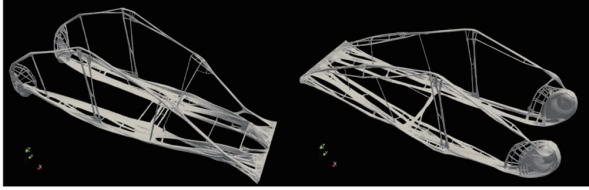
ねじり荷重の場合



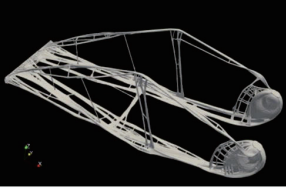
曲げ荷重の場合



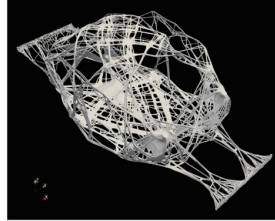
フロントに荷重を作用させた場合



リアに荷重を作用させた場合



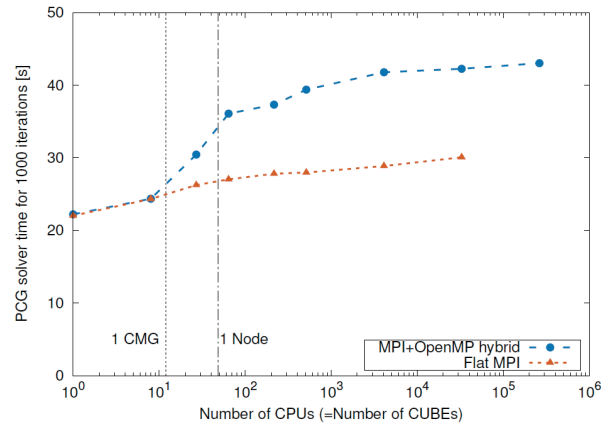
4条件の複合最適化



Computational mesh

- Number of cubes: 34744
- Number of cells: 1.42億
- Minimum cell size: 5 mm

WADA, Yuji, et al. Billion-design-variable-scale topology optimization of vehicle frame structure in multiple-load case. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 2023, 09544070231184309.

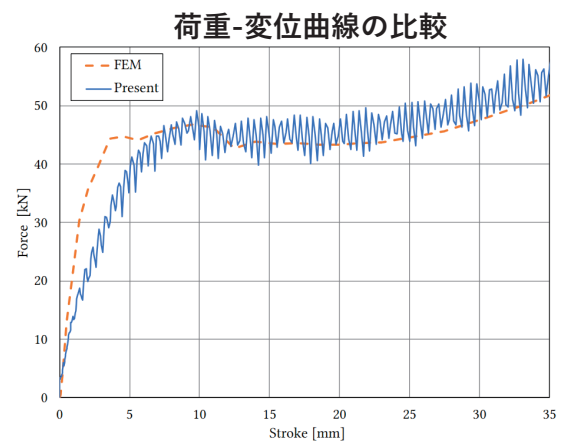
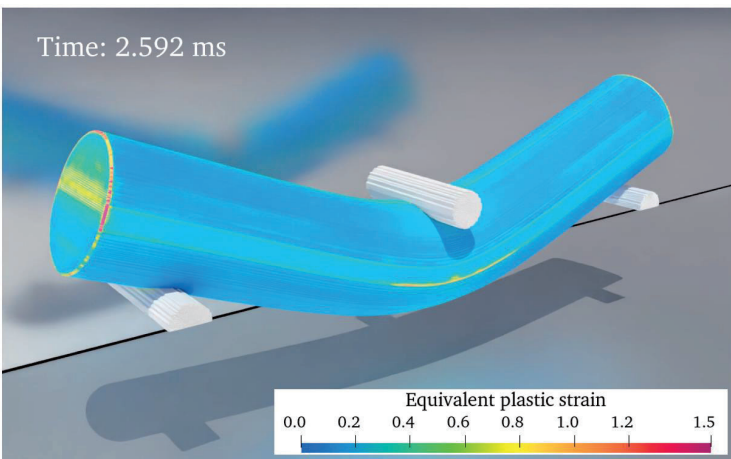
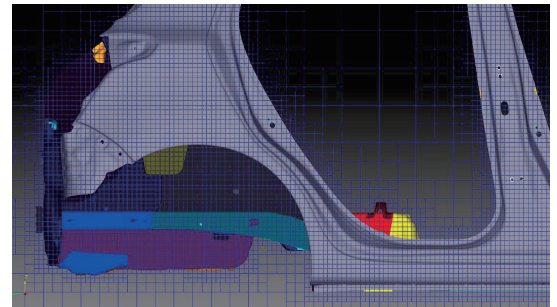
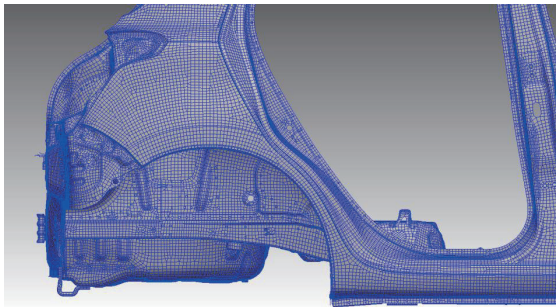


ウィーク・スケーリング@富岳

もくじ

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. **3次元形状のデータセット生成方法**
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - **非線形範囲：オイラー型構造解析**
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. まとめと今後の展望

	Conventional FEM	Eulerian FVM
Mesh generation	lots of time and effort to generate a mesh	Easy and fast to generate a mesh
Large deformation	Mesh distortion can be occurred	Mesh distortion cannot occurred



ウィーク・スケーリング@富岳

Computational mesh

- Number of cubes: 2,560
- Number of cells: 10,485,760
- Minimum cell size: 0.468 mm

Parallel computation

- Hardware: Supercomputer Fugaku
- Node: 80
- MPI Rank: 320
- OpenMP thread: 8
- Total cores: 2560

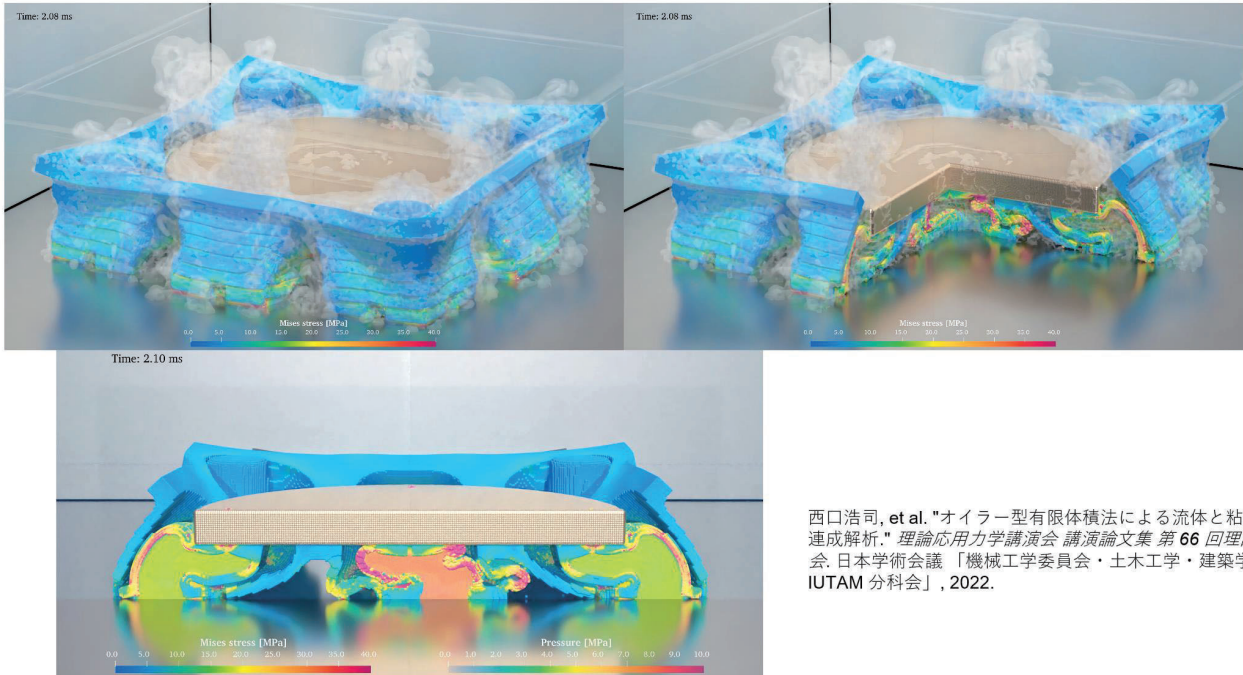
コア数	ノード数	core/node	セル数	実行時間[s]	スケーリング
2048	64	32	8,388,608	17.27	100.0%
16384	512	32	67,108,864	20.13	85.8%
131072	4096	32	536,870,912	23.50	73.5%



幾何学的に複雑な構造の大変形解析が可能

NAGOYA UNIVERSITY

- ゴム系材料のメカニカル・メタマテリアル



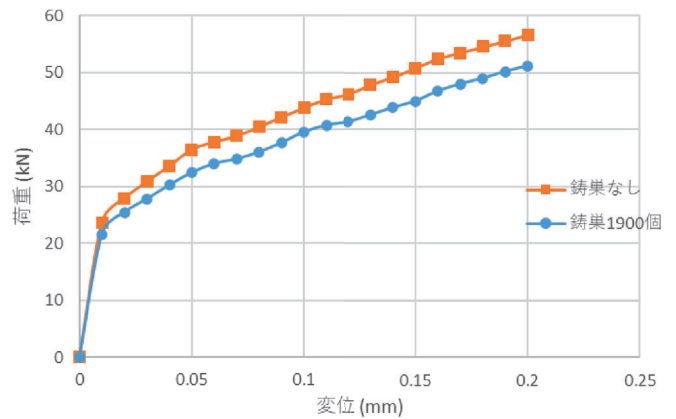
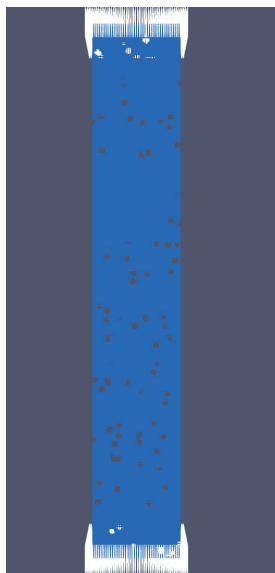
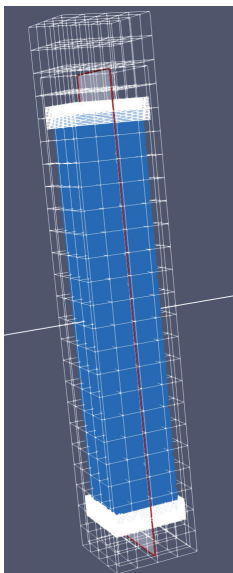
47/74



幾何学的に複雑な構造の大変形解析が可能

NAGOYA UNIVERSITY

- 鋳巣を模擬した試行的解析（弾塑性解析）
- ランダムに鋳巣を設定し、多ケース解析を行うことが可能



- 鋳巣体積率：5%
- 鋳巣の半径：2mm
- 鋳巣の数：約1900個

By 名古屋大学 B4 尾崎 悠利, 2023

48/74



もくじ

NAGOYA UNIVERSITY

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. **3次元形状の深層生成モデル**
7. 試行的な数値計算例
8. まとめと今後の展望

49/74

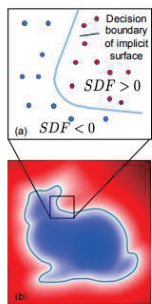


DeepSDF

NAGOYA UNIVERSITY

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, Steven Lovegrove;
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
(CVPR), 2019, pp. 165-174

- 3次元形状をディープニューラルネットワークで陰的に表現
 - ▶ ボクセル, 点群よりメモリ消費が少ない
- 確率的オートデコーダによる3次元形状学習



SDF (距離符号関数)

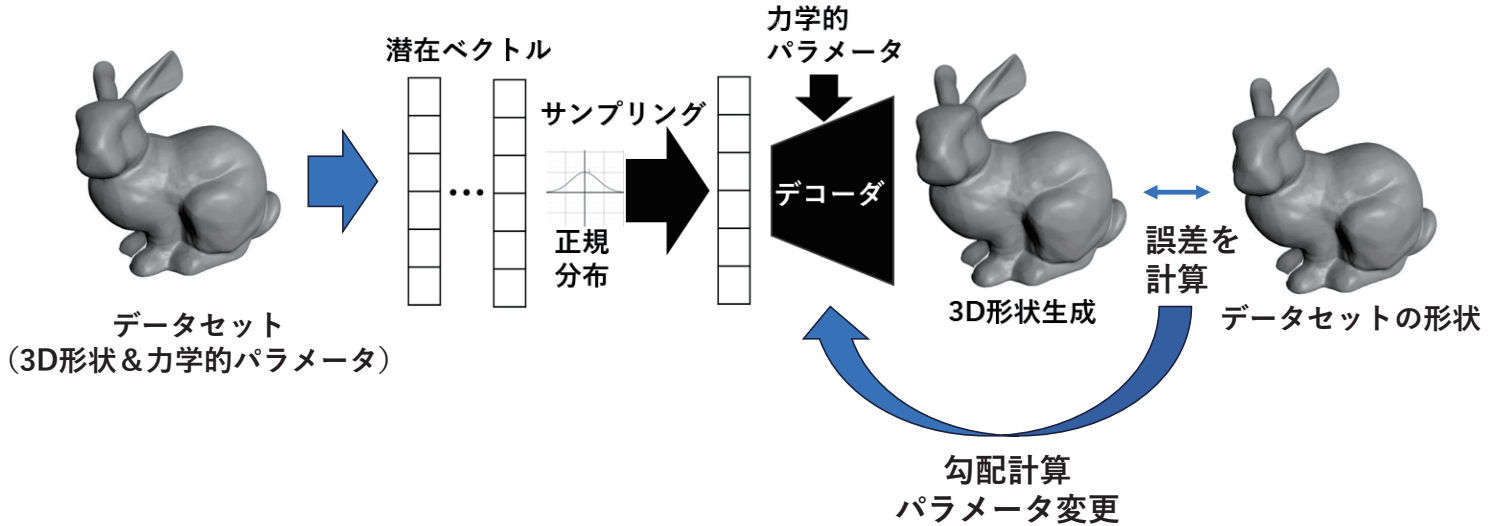
$$x^2 + y^2 = 0$$

$$f(x) = ?$$

50/74

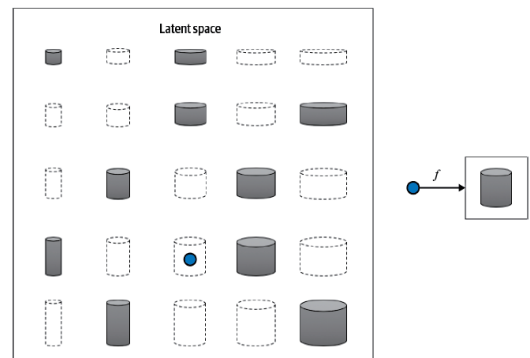
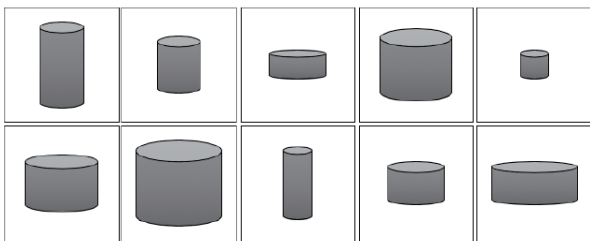
DeepSDFに基づく深層生成モデル

- 力学的パラメータによる条件付き生成が行えるモデル
- 現状のディープニューラルネットは、単純なFFNNを仮定。



表現学習 (Representation Learning)

- 画像/音声/自然言語/3次元形状等のデータから特徴表現 (feature) を自動的に抽出する学習のこと
- 3次元形状の表現学習のイメージ



- 「高さ」と「円の半径」という2つの特徴量で一意に表現できることは人間には明らか。
- しかし、コンピュータには明らかでない。
- 複雑な3次元形状の場合、その形状を一意に表現できる特徴量は、人間の頭では想像できない。

- 3次元形状→潜在空間ベクトル (高さ, 半径)
- 潜在空間ベクトルから3次元形状を写像する関数 f を学習する必要がある。
- 例えば3次元形状のCADデータの形状を変更するために、データをどのように調整すればよいかは自明ではない。
- しかし、潜在空間ベクトルであれば、3次元形状に影響する操作を扱いやすい。



もくじ

NAGOYA UNIVERSITY

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. **試行的な数値計算例**
8. まとめと今後の展望

53/74

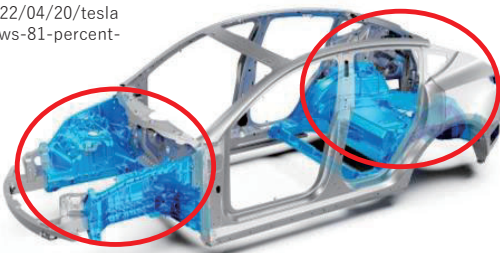


ギガキャスト構造設計の課題

NAGOYA UNIVERSITY

- ① **巨視的スケール：クラッシュブル・ゾーンの最適構造の探索**
 - トポロジー最適化
 - ✓ 線形問題（弾性体の微小変形問題）はほぼ確立されている。
 - ✓ 大変形・破断・弾塑性を考慮できる手法は確立されていない。
- ② **微視的スケール：ランダムに発生する鑄巣の数理モデル化**
 - 応力ひずみ曲線のばらつき
 - 破断ひずみのばらつき

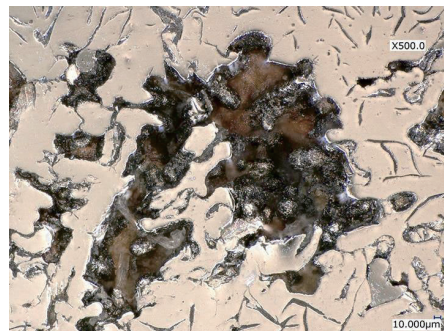
<https://thelastdriverlicenseholder.com/2022/04/20/tesla-revenue-grows-81-percent-in-q1-2022/>



クラッシュブル・ゾーン

(衝突のエネルギーを吸収し、乗員を保護)

<https://www.keyence.co.jp/ss/products/microscope/vhx-casestudy/automobile/die-casting.jsp>



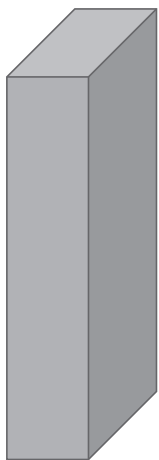
鑄巣

(溶融金属内の空気やガスの混入等による空洞) 54/74

①巨視的スケール：クラッシュブル・ゾーンの最適構造の探索 に対する3次元形状生成AIのフィジビリティ・スタディ

55/74

線形範囲のデータセット生成



初期形状
(設計領域)

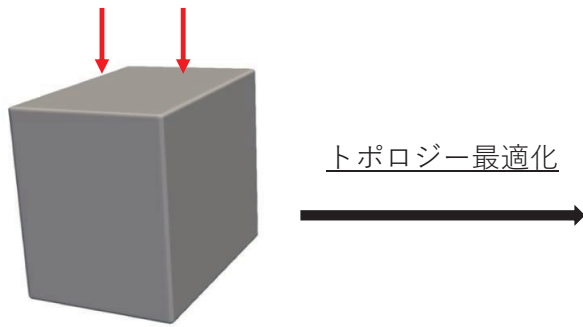
- トポロジー最適化**
- 境界条件(どこを固定するかなど)
 - 荷重条件(どの向きに力がかかるか)
 - 体積条件(初期体積の何%にするか)
 - 目的関数(何を最大化・最小化するか)..ect



形状データ
(訓練データ)

56/74

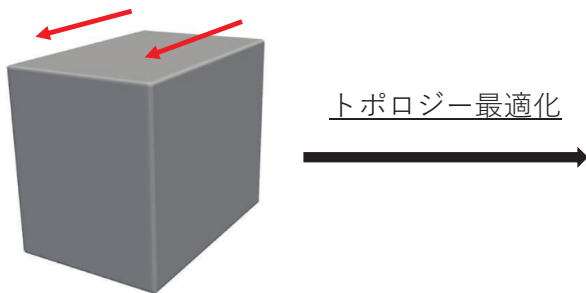
線形範囲のデータセット生成



荷重の向きを変えることで作られる形状が変化



約1000ケース形状データ作成



57/74

計算条件 (スパコン「不老」Type I を利用)

トポロジー最適化

- ◆ Cube数 : 512
- ◆ セル数 : 2,097,152
- ◆ 材料物性値 :
アルミを想定
密度 : 2900kg/m³
ヤング率 : 70GPa
ポアソン比 : 0.3
- ◆ 材料モデル : 弾塑性体
- ◆ トポロジー最適化のステップ数 : 50
- ◆ 実行時間 : 1800s
- ◆ 並列数 : 8 threads × 4 ranks × 16 node
- ◆ 目的関数 :

$$\text{find: } \rho_j, \text{ Minimize: } \Omega = \mathbf{u} \cdot \mathbf{f}$$

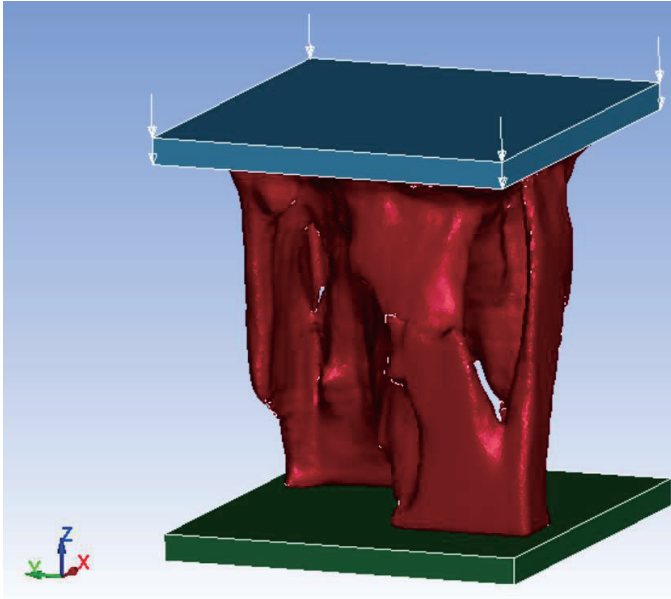
$$\text{Subject to: } \int \rho_j d\Omega_j \leq V_0, \quad \sum_j [k_j] \mathbf{u} = \mathbf{f}$$

弾塑性解析

- ◆ Cube数 : 512
- ◆ セル数 : 2,097,152
- ◆ 材料物性値 :
アルミを想定
密度 : 2900kg/m³
ヤング率 : 70GPa
ポアソン比 : 0.3
- ◆ 材料モデル : 弾塑性体
- ◆ 順解析のステップ数 : 1
- ◆ 実行時間 : 10800s
- ◆ 並列数 : 8 threads × 4 ranks × 16 node
- ◆ 荷重条件 : zマイナス方向(圧縮)
- ◆ 境界条件 : トポロジー最適化と同一

58/74

非線形範囲のデータセット生成



- ◆ Cube数 : 512
- ◆ セル数 : 2,097,152

◆ 材料物性値 :

アルミを想定

密度 : 2900kg/m³

ヤング率 : 70GPa

ポアソン比 : 0.3

- ◆ 材料モデル : 弾塑性体
- ◆ 順解析のステップ数 : 1
- ◆ 実行時間 : 10800s
- ◆ 並列数 : 8 threads × 4 ranks × 16 node
- ◆ 荷重条件 : zマイナス方向(圧縮)
- ◆ 境界条件 : トポロジー最適化と同一

59/74

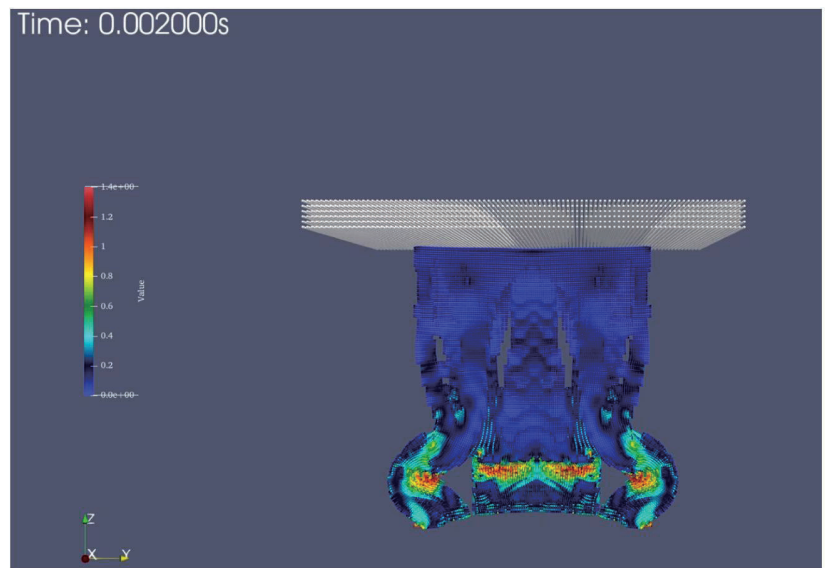
非線形範囲のデータセット生成

線形範囲での力学的パラメータ

- 質量
- 体積
- 剛性(線形範囲)
- ひずみエネルギー(線形範囲)..ect

非線形範囲での力学的パラメータ

- 強度(非線形範囲)
- ひずみエネルギー(非線形範囲)..ect

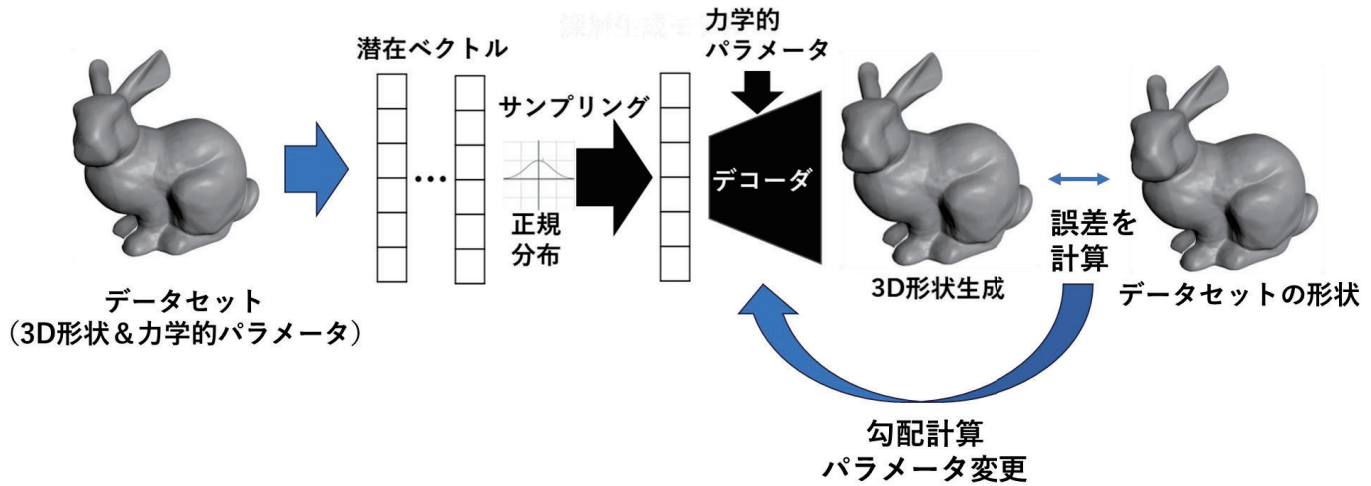


秒速20m/sで衝突解析
ひずみ(変形量)を可視化

60/74

DeepSDFに基づく深層生成モデルの訓練

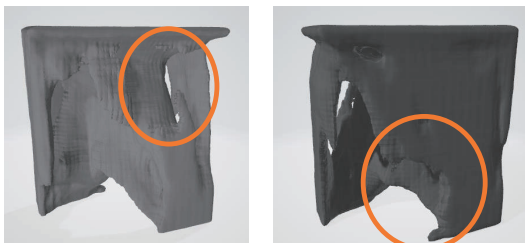
- 1092ケースの3次元形状と力学的パラメータのペアデータセットを用いて、深層生成モデルを訓練した
 - 名古屋大学 スパコン「不老」 type II (V100 GPU : 1/4ノード)
 - 500epoch, 約6時間



3次元形状の生成例

指定した力学的パラメータ：ひずみエネルギー ($N \cdot m^2$)

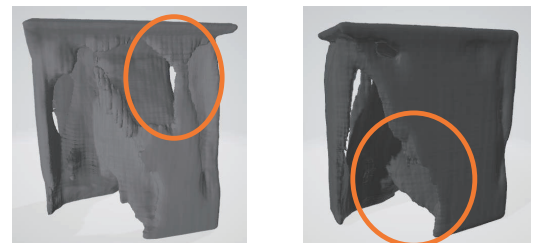
ひずみエネルギー： 1.57×10^{-10}



正面

背面

ひずみエネルギー： 5.25×10^{-9}



正面

背面

- ひずみエネルギーを指定することで、3次元形状を生成できることを確認
- ひずみエネルギーを増加させると、ひずみを受け持つ構造が追加され（上図○領域）、定性的に力学的合理性のある3次元形状の生成が行えることを確認した。



NAGOYA UNIVERSITY

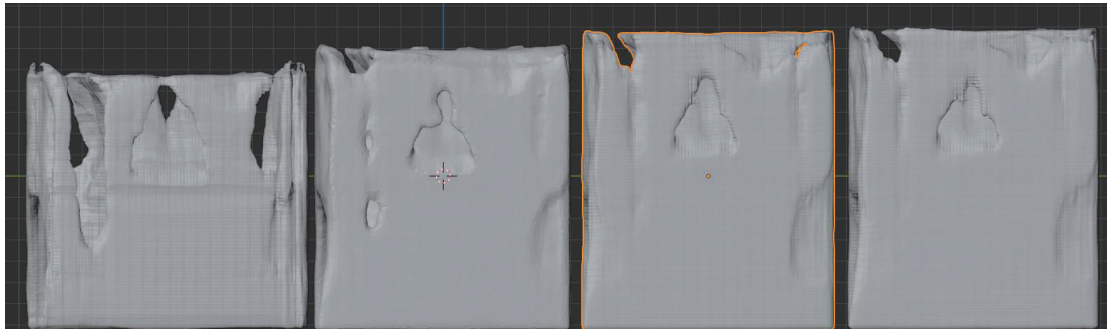
生成された3次元形状の精度検証

力学的パラメータ：高さ(mm)

• 訓練データ数： $\frac{52}{\text{荷重方向}} \times \frac{21}{\text{初期形状}} = 1092$

真値(入力値)	100mm	130mm	150mm	160mm
---------	-------	-------	-------	-------

再構築形状



高さ

予測値(寸法)	121mm	131mm	141mm	142mm
---------	-------	-------	-------	-------

予測精度	79%	99%	94%	89%
------	-----	-----	-----	-----

63/74



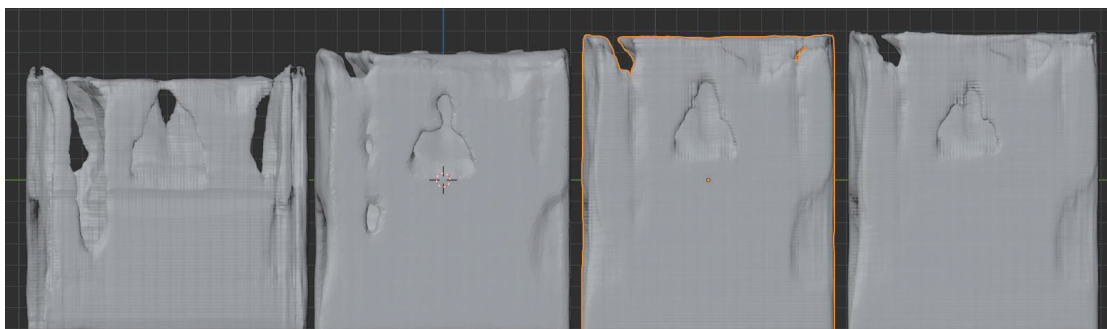
NAGOYA UNIVERSITY

生成された3次元形状の精度検証

力学的パラメータ：ひずみエネルギー(N・m²)

真値(入力値)	4.32×10^{-8}	4.32×10^{-8}	4.32×10^{-8}	4.32×10^{-8}
---------	-----------------------	-----------------------	-----------------------	-----------------------

再構築形状



予測値(寸法)	3.83×10^{-9}	3.28×10^{-9}	4.17×10^{-9}	5.04×10^{-9}
---------	-----------------------	-----------------------	-----------------------	-----------------------

予測精度	8.7%	7.6%	9.6%	12.5%
------	------	------	------	-------

64/74



試行的取り組みのまとめ

• まとめ

- ▶ DeepSDFに基づく深層生成モデルで、力学的パラメータから3次元形状を生成できる見通しが立った。
- ▶ ただし、現在の試行的計算では、データの数が十分でないため、指定した力学的パラメータを満足しない3次元形状が生成される。

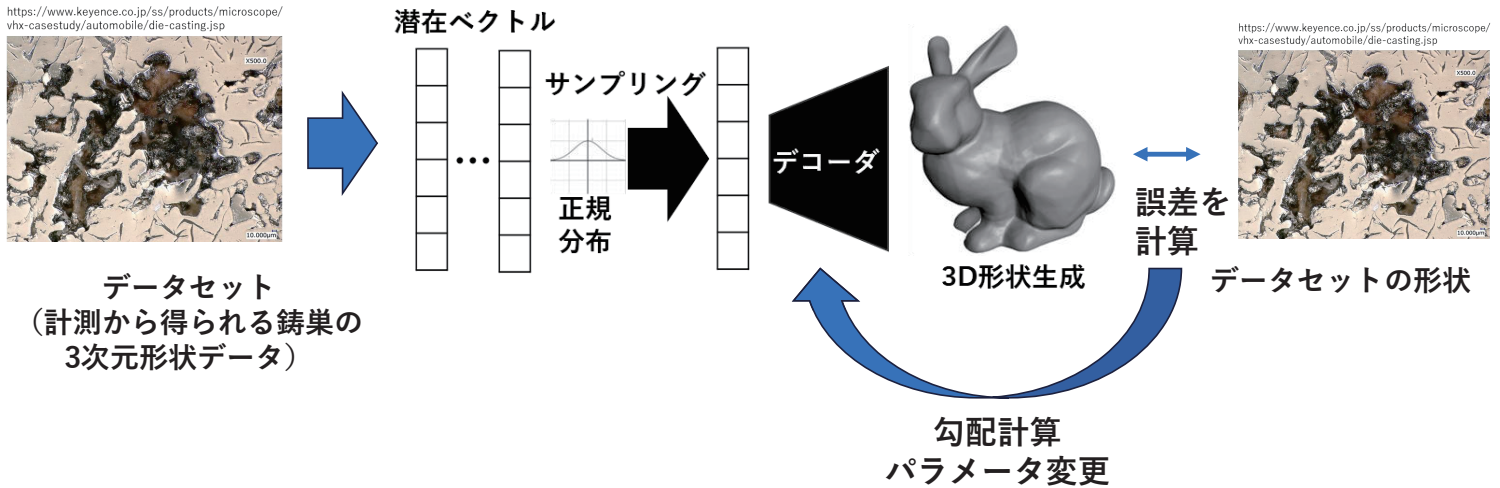
• 今後の予定

- ▶ 衝撃吸収構造部材に特化し、1万個を目途にデータセットを生成
- ▶ DNNアーキテクチャの差異の調査（FFNN, Transformer等）
- ▶ 力学的パラメータから生成された3次元形状の検証
 - 「富岳」一般課題（重点課題）「次世代自動車設計のためのコンポーネント構造の超多ケース弾塑性解析と確率的深層生成モデルへの展開」（課題番号：hp230296, 研究課題代表者：西口浩司, 11,687,500ノード時間, 令和5年度下期～令和6年度上期）の計算資源で推進予定



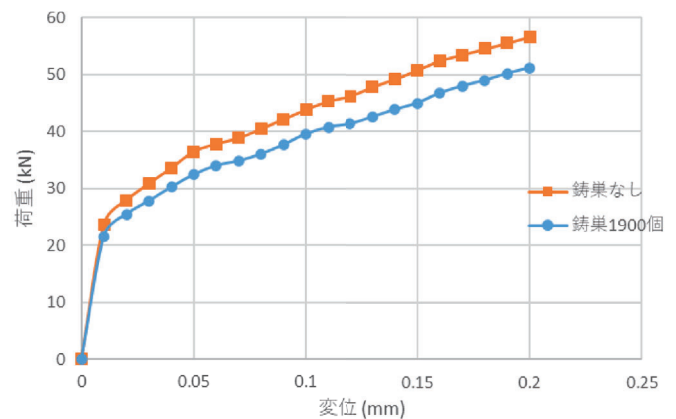
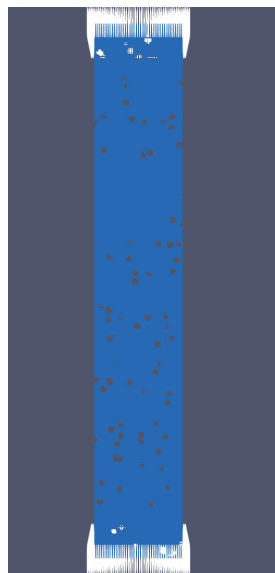
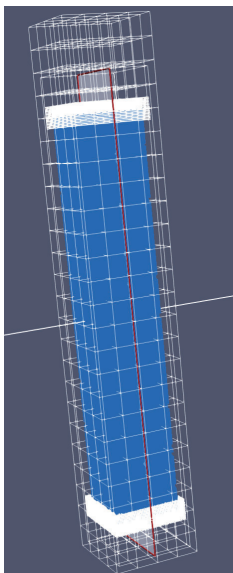
微視的スケール：ランダムに発生する鋳巣の数理モデル化 に対する3次元形状生成AIのフィジビリティ・スタディ

- DeepSDFにランダムに発生する鋳巣の3次元形状を学習させる



67/74

- DeepSDFにより生成した3次元形状を鋳巣形状としてモデル化
- オイラー型弾塑性解析を行い、多ケースの応力ひずみ曲線を取得



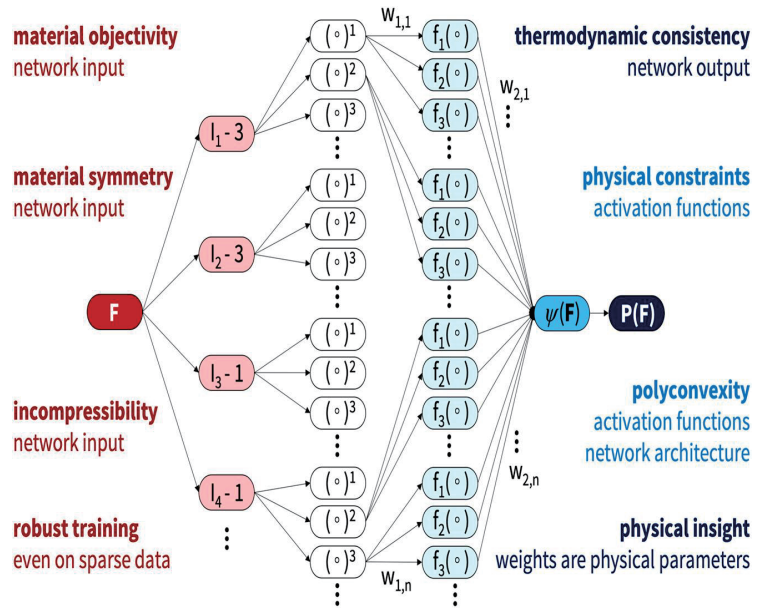
- 鋳巣体積率：5%
- 鋳巣の半径：2mm
- 鋳巣の数：約1900個

By 名古屋大学 B4 尾崎 悠利, 2023

68/74

NNによる構成則（応力ひずみ関係）のモデル化

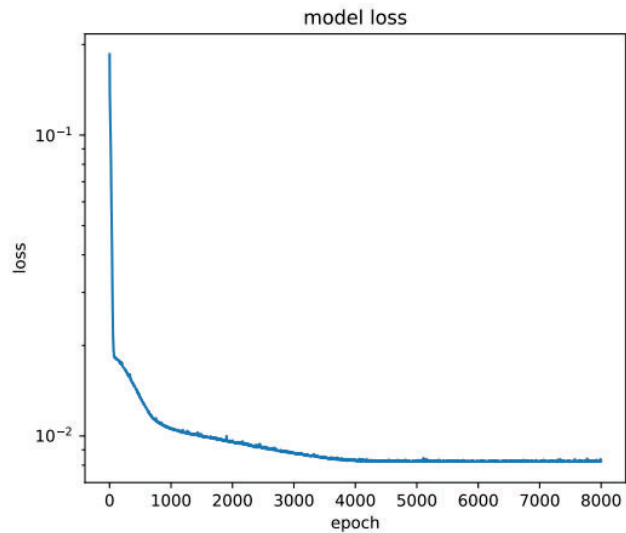
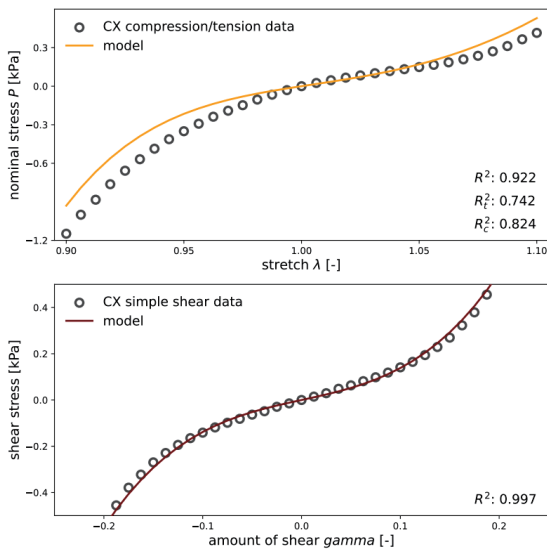
- 物質の構成方程式をNNでモデル化すると、一般に物質客観性の原理を満足しない。
- スタンフォード大学の研究グループが提案したConstitutive Artificial Neural Networkは、超弾性構成方程式を一般化したものであるため、物質客観性の原理を満足する。



Linka, K., & Kuhl, E. (2023). A new family of Constitutive Artificial Neural Networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403, 115731.

試行的計算例

- ひずみエネルギー関数をより正確に同定するには、圧縮・引張・せん断を同時に考慮することが必要
- 圧縮・引張・せん断を同時に学習できることを確認





試行的取り組みのまとめ

- オイラー型弾塑性解析により，鋳巣形状をランダムに発生させたアルミニウム構造の応力ひずみ曲線を算出できる見通しが立った。
- DeepSDFに基づく深層生成モデルで，ランダムに発生する鋳巣の3次元形状をモデル化した上で，オイラー型構造解析により多ケースの応力ひずみ曲線のデータセットを生成する予定。
- 応力ひずみ曲線のばらつきや破断ひずみのばらつきは，スタンフォード大学の研究グループのNNに確率層を導入することでモデル化する予定。

71/74



もくじ

1. 3次元形状生成AIの急速な発展
2. 2つのスケーリング則
3. スーパーコンピュータ「富岳」と第二階層スパコン
4. ギガキャスト構造の利点と設計上の課題
5. 3次元形状のデータセット生成方法
 - 階層型の固定直交メッシュ法
 - 線形範囲：トポロジー最適化
 - 非線形範囲：オイラー型構造解析
6. 3次元形状の深層生成モデル
7. 試行的な数値計算例
8. **まとめと今後の展望**

72/74



まとめと今後の展望

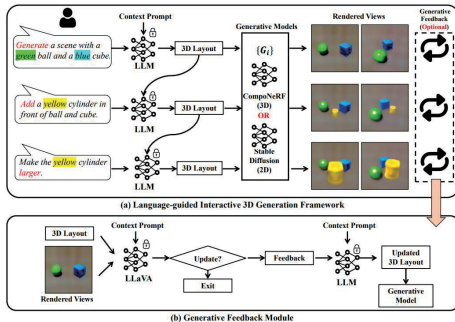
NAGOYA UNIVERSITY

• まとめ

- 3次元形状生成AIが急速に発展している技術的背景を概説しました。
- 次世代自動車構造として注目されているギガキャスト構造の最適設計を目指した3次元形状生成AI構築の試行的取り組み（フィジビリティ・スタディ）をご紹介しました。

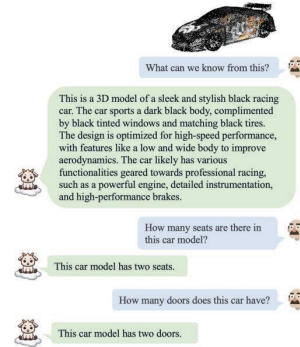
• 今後の展望

- 3次元形状生成モデルの**大規模化（GPT-FugakuにおけるTransformer）**による性能向上の可能性探求。LLM同様に創発性は観測されるのか？
- CAEの民主化：**大規模言語モデル（GPT-Fugaku）**による力学的パラメータの制御および数値解析結果（3次元データ）解釈の可能性探求。



LLMs as Layout Interpreter

<https://arxiv.org/pdf/2305.15808.pdf>



LLMs to understand 3D Point Clouds

<https://arxiv.org/pdf/2308.16911.pdf>

73/74



謝辞

NAGOYA UNIVERSITY

本発表の数値計算は、以下の計算資源により実施しました。

ここに記して感謝申し上げます。

- 令和4年度B期「富岳」利用研究課題（課題番号：hp22024，次世代自動車設計のためのコンポーネント構造の大変形・破断を伴う弾塑性解析，2022年10月～2023年9月，代表者：西口浩司）
- JSPS科研費22H00226により購入したスーパーコンピュータ「不老」の計算資源

74/74

文部科学省 原子力システム研究開発事業(ボトルネック課題解決型)

AI技術を活用した 確率論的リスク評価手法の高度化研究

全体計画と弊社の取組み紹介

第5事業部 主管研究員 森本 達也

【技術セミナー】生成AI・大規模言語モデルとCAE

2023年9月22日(金)

アドバンスソフト株式会社

紹介内容

1. 全体計画
 - (1) 目的
 - (2) 背景
 - (3) 体制
補足: 原子力発電所のPRA
 - (4) 開発対象
2. 弊社の取組み
 - (1) 開発計画・背景
 - (2) NUCIA用の自動故障判定手法の開発
 - (3) CORDS用の自動故障判定手法の開発
3. 終わりに

紹介内容

1. 全体計画

(1) 目的

(2) 背景

(3) 体制

補足：原子力発電所のPRA

(4) 開発対象

2. 弊社の取組み

(1) 開発計画・背景

(2) NUCIA用の自動故障判定手法の開発

(3) CORDS用の自動故障判定手法の開発

3. 終わりに

(1) 目的

課題とその要因：

- 原子力発電所の確率論的リスク評価(PRA: Probabilistic Risk Assessment)は、解析作業が膨大で事業者の大きな負担。
- 膨大な設計資料等の読み込みと理解、信頼性データや評価モデルの構築を、習熟した技能者が経験に基づいて手作業で入力しなければならない状況。



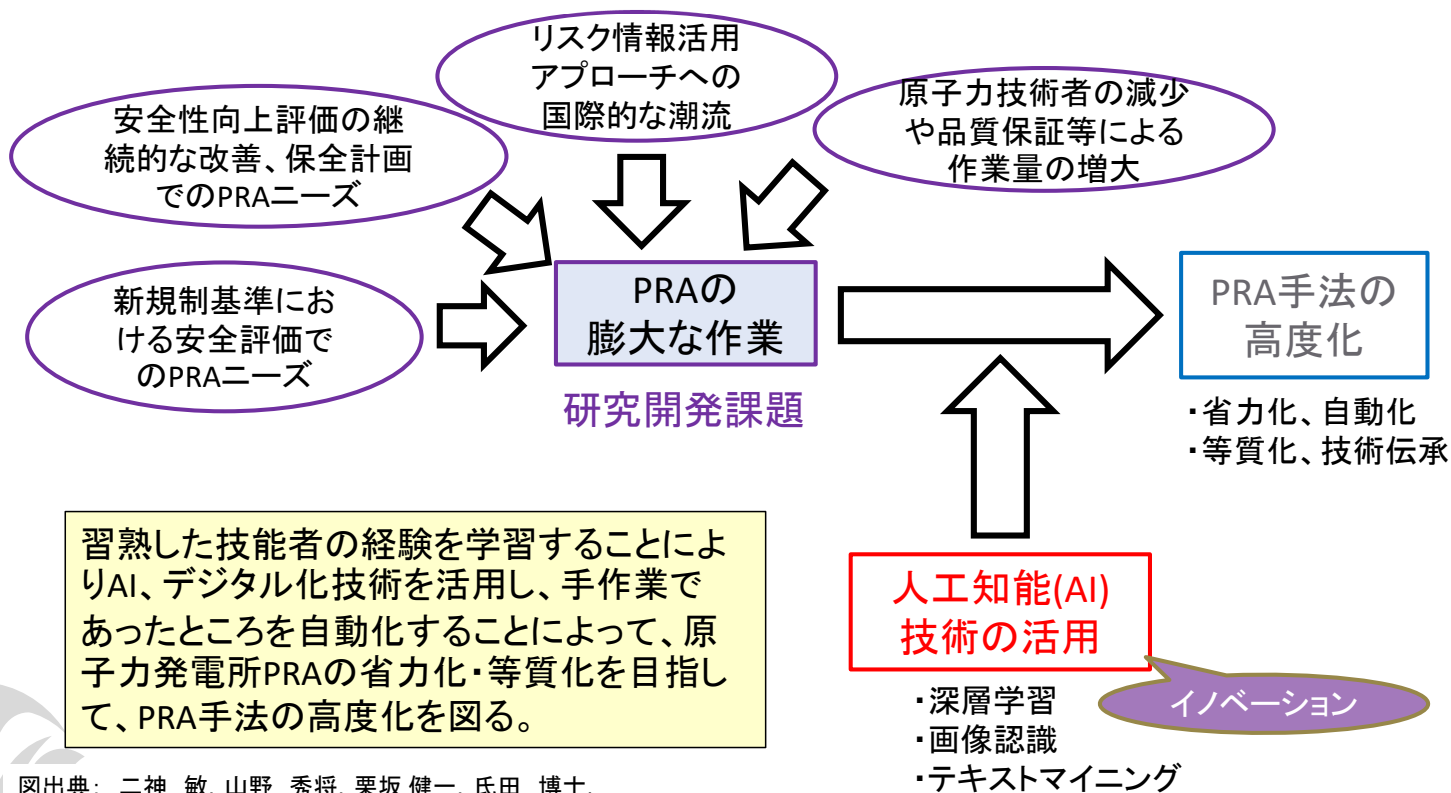
AI、デジタル化技術を活用し、
手作業であったところを自動化することによって、
原子力発電所**PRAの省力化・等質化**を目指す。

本研究では、原子力発電所運転時のPRAにおける

「①フォルトツリー自動作成手法の開発」及び

「②信頼性データベース構築のための自動故障判定手法の開発」に着目し、
2022年度からの3か年でAIツールを開発し、PRA手法の高度化を図る。

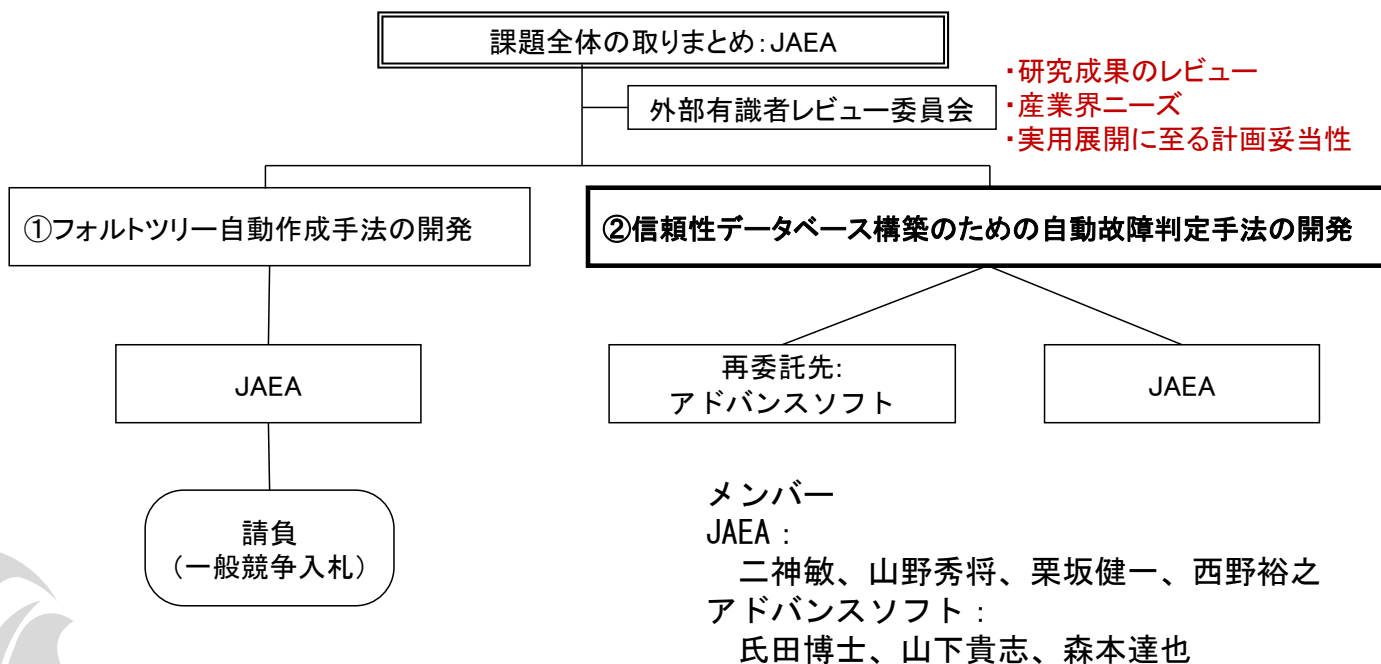
(2) 背景



図出典：二神 敏, 山野 秀将, 栗坂 健一, 氏田 博士, 「AI技術を活用した確率論的リスク評価手法の高度化研究(1)AIツールの開発計画」日本原子力学会 2023年春の年会

(3) 体制

文部科学省からの委託業務として、国立研究開発法人 日本原子力研究開発機構 (JAEA)と共に、以下の②を弊社(アドバンスソフト)が実施中。



紹介内容

1. 全体計画

(1) 目的

(2) 背景

(3) 体制

補足：原子力発電所のPRA

(4) 開発対象

2. 弊社の取組み

(1) 開発計画・背景

(2) NUCIA用の自動故障判定手法の開発

(3) CORDS用の自動故障判定手法の開発

3. 終わりに

補足 原子力発電所のPRA

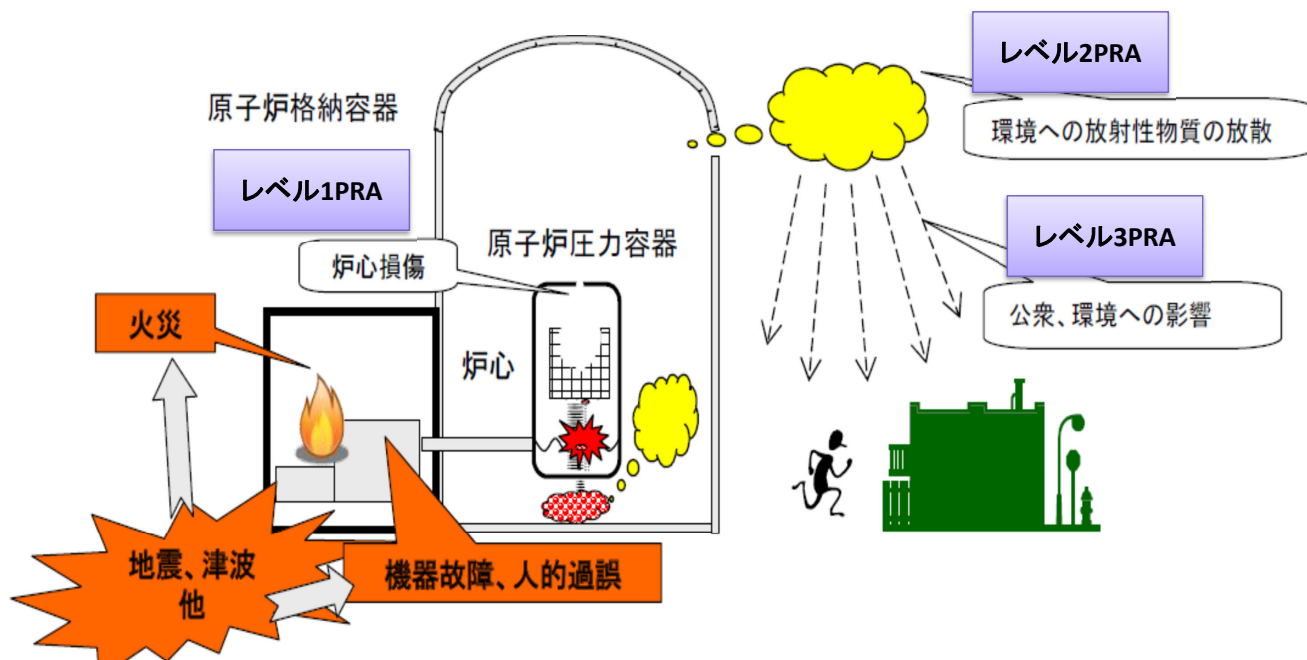
原子力発電所の安全性に関するリスクを、確率論を用いて総合的かつ定量的に評価するもの。

以下に示すような多岐にわたる領域を包括的に分析する。

- **対象施設：** 発電所全体、関連施設
- **施設の運転状態：** 運転時、停止時
- **発生事象の範囲：** 内的事象（施設内部で起こる事象）、外的事象（自然災害など）
- **評価指標の範囲：**
 - レベル1・・・炉心損傷又は燃料損傷の発生頻度までを評価
 - レベル2・・・格納容器破損の発生頻度及びその場合の放射性物質の環境放出量等までを評価
 - レベル3・・・公衆や環境への影響の発生頻度と大きさまでを評価
- **主な分析手法**
 - イベントツリー(ET)分析: 事象の進行を分析し、各事象シナリオの確率を計算
 - フォールトツリー(FT)分析: 各システムの故障モードを分析し、それぞれの故障の確率を計算

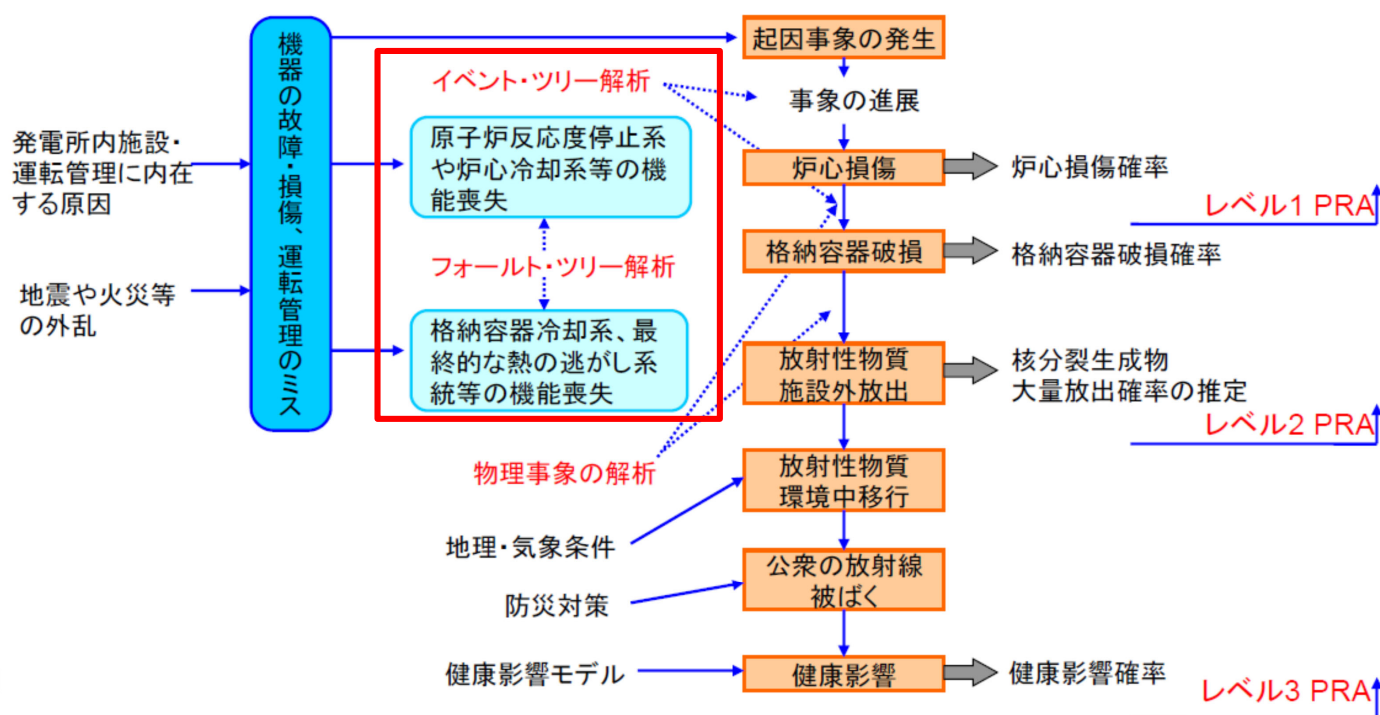
これにより、対象施設の安全対策を最適化し、より効果的なリスク管理を実現。

補足 原子力発電所のPRA



図出典:「原子力プラントの包括的なリスク評価」 成宮祥介、日本原子力学会2013秋の大会 講演資料

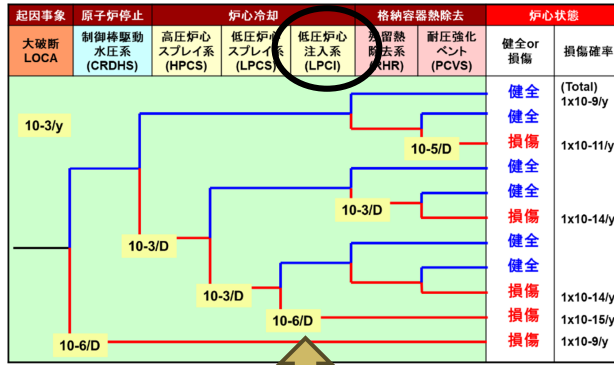
補足 原子力発電所のPRA



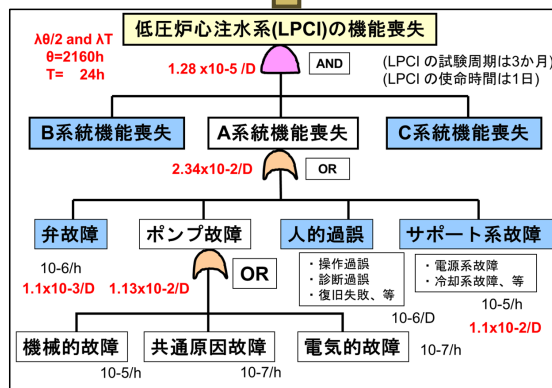
図出典:「各国(日本、米国、英国、仏国)における確率論的リスク評価の活用状況」 平成27年1月 経済産業省

補足 原子力発電所のPRA

イベントツリー
ET



フォルトツリー
FT



①フォルトツリー自動作成手法の開発

設計図書(系統図)
機器仕様図書など

信頼性データ
(故障発生個所、故障モード、原因など)

②信頼性データベース構築のための自動故障判定手法の開発

図出典: 氏田博士、柚原直弘:『システム安全学-文理融合の新たな専門知』2015.9、海文堂出版

紹介内容

1. 全体計画

- (1) 目的
- (2) 背景
- (3) 体制

補足: 原子力発電所のPRA

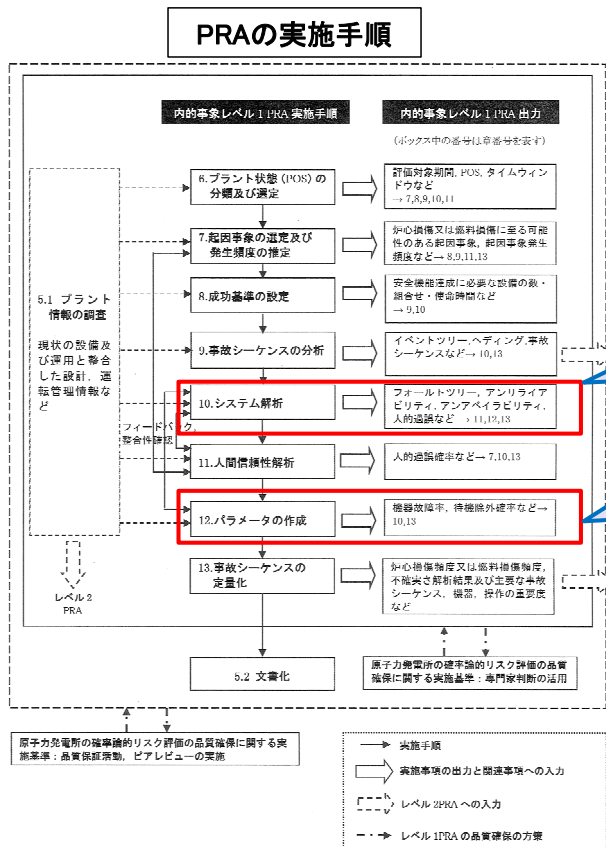
(4) 開発対象

2. 弊社の取組み

- (1) 開発計画・背景
- (2) NUCIA用の自動故障判定手法の開発
- (3) CORDS用の自動故障判定手法の開発

3. 終わりに

(4) 開発対象



作業量が多く、省力化の効果、AI技術の適用を期待

①フォルトツリー自動作成手法の開発

②信頼性データベース構築のための自動故障判定手法の開発

“原子力発電所の内的事象を起因とした確率論的リスク評価に関する指針(レベル1 PRA 編)：2022”より抜粋

(4) 開発対象

②信頼性データベース構築のための自動故障判定手法の開発

目標：
テキストマイニングやデータマイニング等のAI技術を活用し、各原子力プラントの故障及びトラブル情報から、PRAに必要な故障を自動的に判定する手法を開発し、信頼性データベース構築に資する。

詳細は「2. 弊社の取組み」で。

紹介内容

1. 全体計画

(1) 目的

(2) 背景

(3) 体制

補足：原子力発電所のPRA

(4) 開発対象

2. 弊社の取組み

(1) 開発計画・背景

(2) NUCIA用の自動故障判定手法の開発

(3) CORDS用の自動故障判定手法の開発

3. 終わりに

(1) 開発計画・背景

■ 2022年度

- NUCIA※¹から、信頼性データベース構築に必要な情報を抽出してデータベース化する方法論を構築し、ツールを試作。

■ 2023年度

- CORDS※²に基づいて、事象データの自動故障判定ツールを試作。
- 信頼性データの事象や時間の相違を分析することにより、電力会社、プラント及び職種の相違判断、福島第一原子力発電所事故前後の相違判断等が可能な手法を試作中。

■ 2024年度

- 共通特性を分析することにより、共通機器、共通操作、共通組織特性等の共通要因を判断可能な手法を試作予定。

- ※1 NUCIA (<http://www.nucia.jp/>) : 国内原子力発電所等におけるトラブル情報等が登録・公開されているサイト
- ※2 CORDS: JAEAが整備されている高速炉用の機器信頼性データベース

(2) NUCIA用の自動故障判定手法の開発

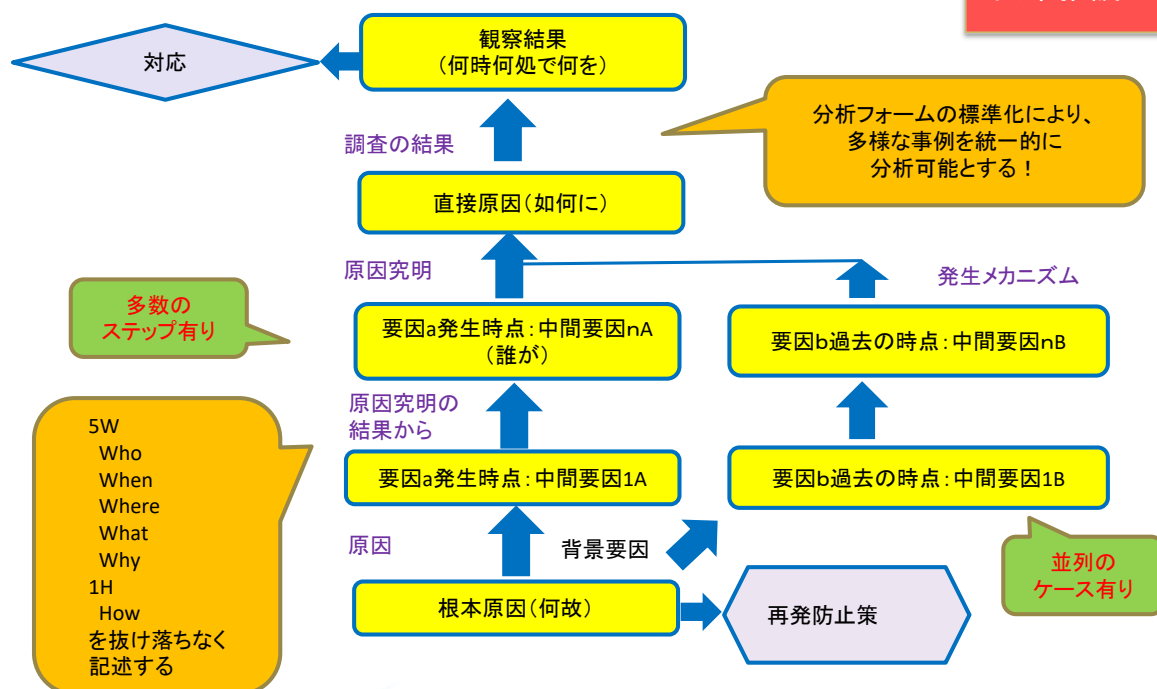
NUCIAにおけるトラブル情報等の報告書の特徴と、それへの対応策

- 各電力会社の各発電プラントの各担当者から報告されており、項目は統一されているが記述の個別性が高く、信頼性データベース化するには不要な情報が多い。
⇒まずは原情報を取り込み、不要情報の削除が必要。
- 記述の個別性として、トラブル情報によって記述形式や記述量に大きな違いがある。
⇒記述形式・量の統一のため、テキストクリーニング、文章要約、キーフレーズ抽出による整理が必要。
- 故障判定には、原因の特定(原因分類)が必要。既存の原因分類は、1つのトラブル情報に対して1つの原因のみが選定されているケースが多いが、本当にそうか？また、そもそも既存の分類項目は適切か？
⇒適切な分類項目に基づいたマルチラベル分類が必要。
- 原因の特定に必要な情報が本当に網羅されているか？欠損情報を補間できないか？
⇒根本原因から事象発生、再発防止策に至るまでの事象シナリオを構築し、ネットワーク図として可視化することで、情報欠落の確認や説明性の高い原因特定が可能と考える。また、事象や時間の相違を分析しやすいはず。

(2) NUCIA用の自動故障判定手法の開発

試作している事象シナリオのネットワーク図のイメージ オントロジー(Ontology、情報表現の仕様)

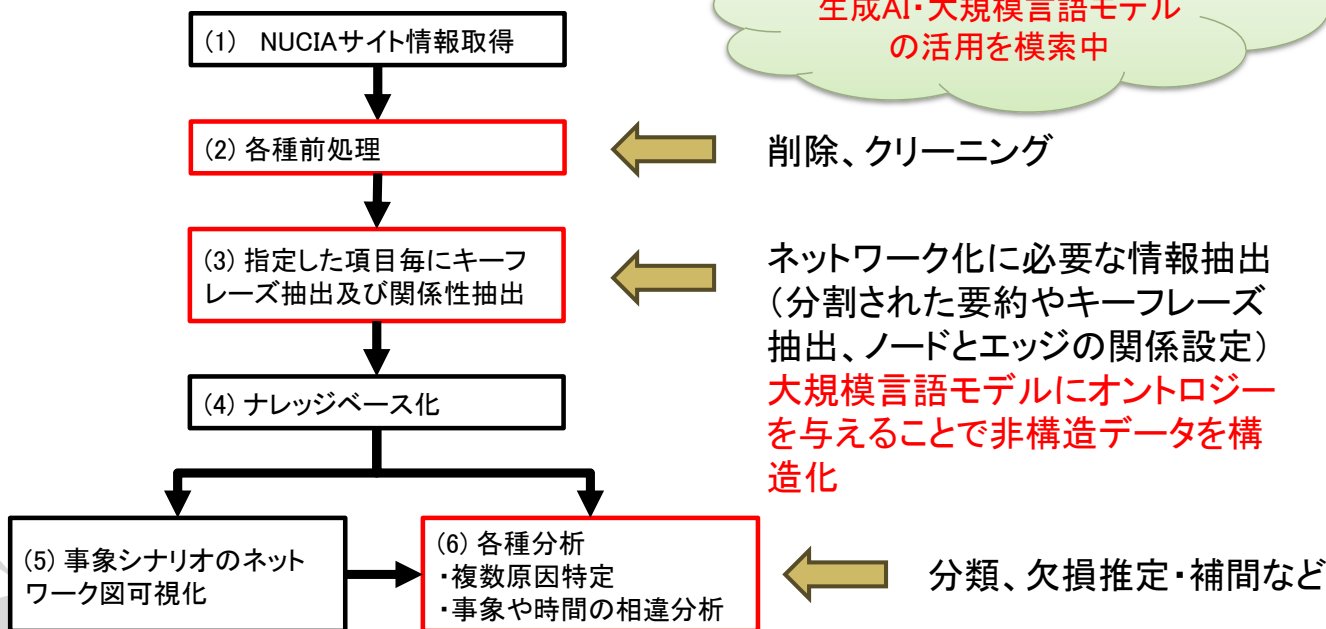
因果推論が重要！



(2) NUCIA用の自動故障判定手法の開発

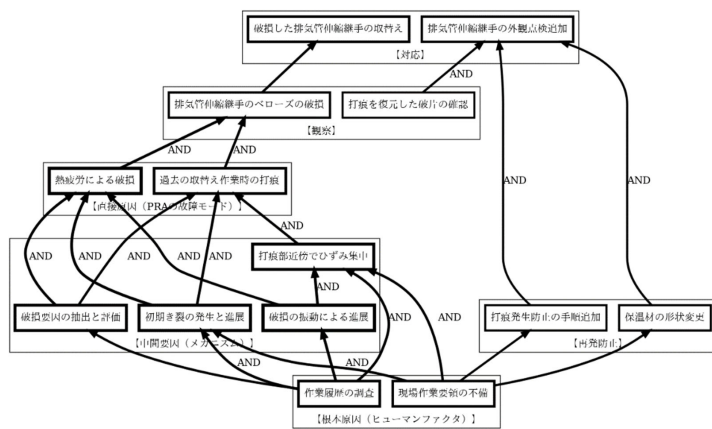
2022年度に試作した要素技術を踏まえつつ、2023年度版を試作中。

赤枠部分：
生成AI・大規模言語モデル
の活用を模索中



(2) NUCIA用の自動故障判定手法の開発

現状の課題



- 教師データが無い及び十分な量を作成できないため、抽出結果や接続結果の精度及び統一性に問題あり。
- 効率的・効果的な妥当性検証方法を要検討。弊社のリスク評価専門家がいくつかピックアップしてチェックしているのが現状。

紹介内容

1. 全体計画

- (1) 目的
- (2) 背景
- (3) 体制

補足：原子力発電所のPRA

- (4) 開発対象

2. 弊社の取組み

- (1) 開発計画・背景
- (2) NUCIA用の自動故障判定手法の開発
- (3) CORDS用の自動故障判定手法の開発

3. 終わりに

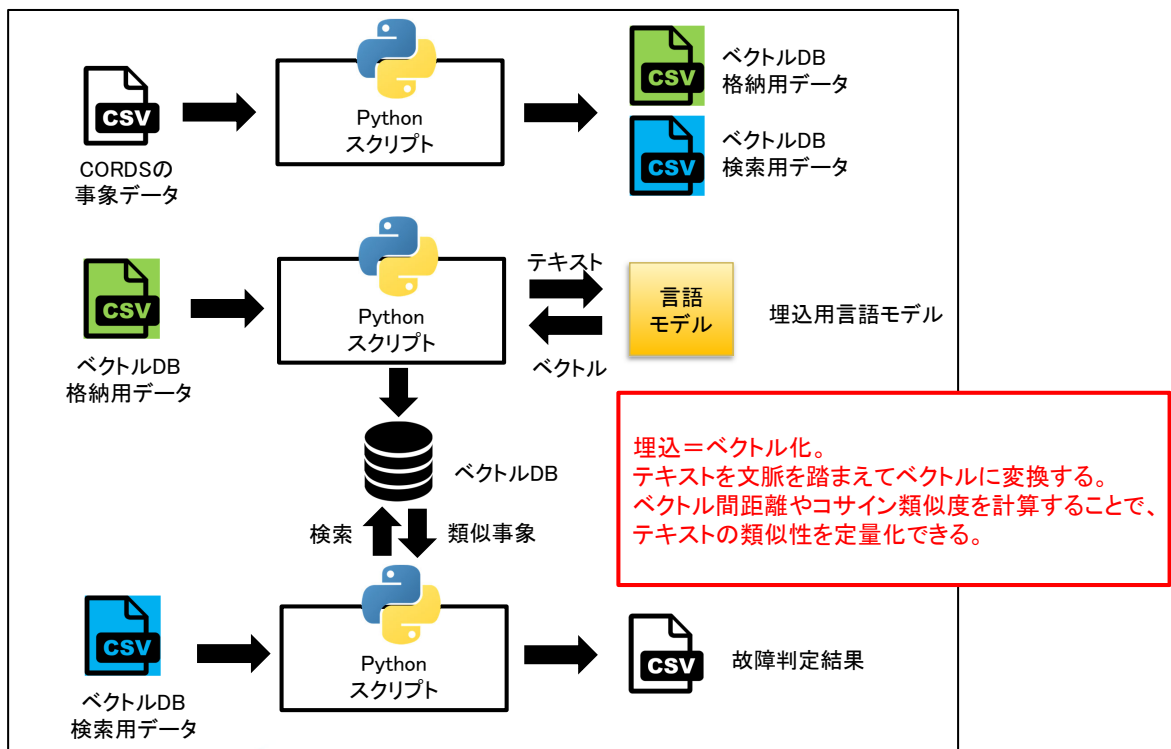
(3) CORDS用の自動故障判定手法の開発

CORDSにおけるトラブル情報等の報告書の特徴と、それへの対応策

- トラブル情報の分析件数、分析対象、分析者が限定されており、各事例の粒度が均一であり、事象データとそれに対する故障判定結果が整理されている。
⇒出来る限り、過去の事例に基づくルールベースな手法が適切であると判断
- JAEAと共同で定めた今年度の開発目標：
事象データを入力すれば、当該事象について以下の判定結果を出力する。
 - ✓ 機器バウンダリ内か否かの判定結果
 - ✓ 機器バウンダリ内故障の場合には事象の程度が機能喪失か否かの判定結果
 - ✓ 機能喪失以外の場合には損傷の程度が甚大か否かの判定結果
 - ✓ 機能喪失又は甚大との判定結果が得られた場合にはどの故障モードに該当するかの判定結果
 ⇒これらの判定はテキスト分類処理で対応可能

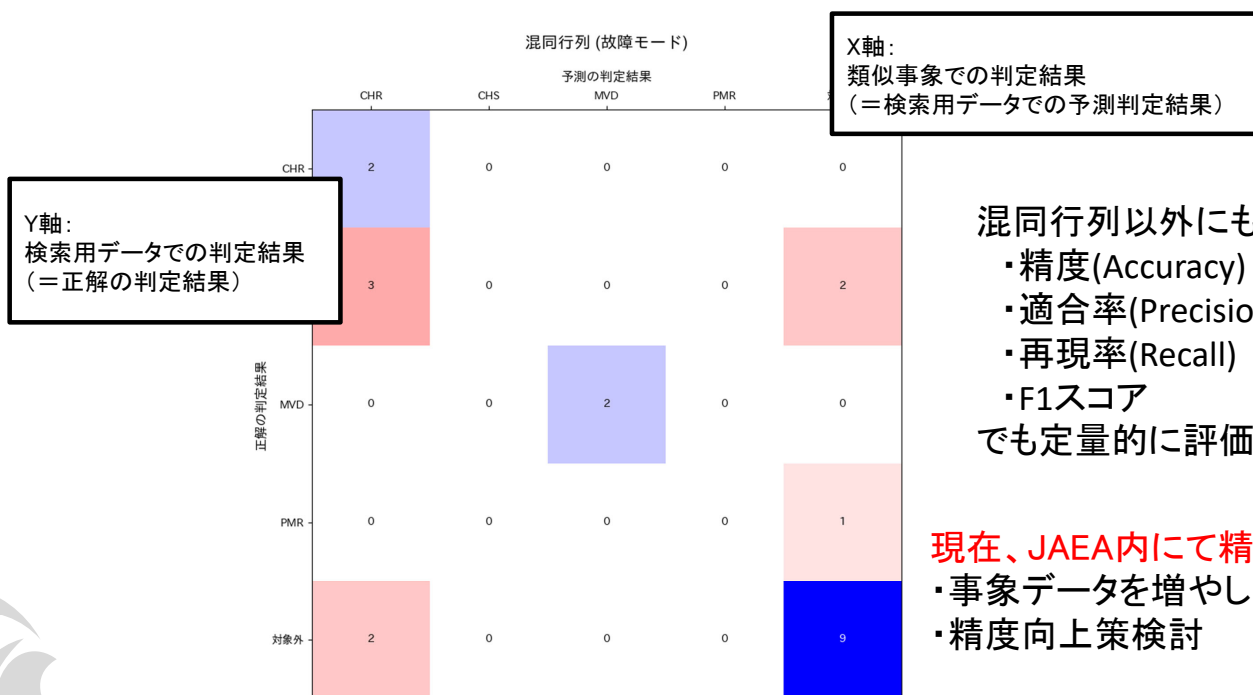
(3) CORDS用の自動故障判定手法の開発

試作したツールの処理フロー全体イメージ



(3) CORDS用の自動故障判定手法の開発

試作したツールによる処理結果例(格納用82、検索用21件)



紹介内容

1. 全体計画

(1) 目的

(2) 背景

(3) 体制

補足：原子力発電所のPRA

(4) 開発対象

2. 弊社の取組み

(1) 開発計画・背景

(2) NUCIA用の自動故障判定手法の開発

(3) CORDS用の自動故障判定手法の開発

3. 終わりに

3. 終わりに

- 原子力発電所のPRAの省力化・等質化を目指して、自動化ツールを試作中。
- 各原子力発電所のトラブル情報等のテキスト(非構造データ)を構造化するために、生成AI・大規模言語モデルの活用を模索中。
- その有用性を確認しつつあるが、生成AI・大規模言語モデルはあくまで選択肢の1つであり、本当に課題解決につながるかは試行錯誤中。
- オープンソース、高性能、日本語対応なモデルがあれば、今後も試用する予定。より良い実用的な結果を得るために。
- CAEにおける利用のポイントは、非構造データの構造化だと考えられる。

生成AI導入支援サービスの紹介

第5事業部 主管研究員 佐藤 智昭

【技術セミナー】生成AI・大規模言語モデルとCAE
2023年9月22日（金）
アドバンスソフト株式会社

アジェンダ

1. 背景
2. 生成AIの概要
3. 導入支援サービス
4. おわりに

背景: AI(人工知能)にかかわる技術の進展

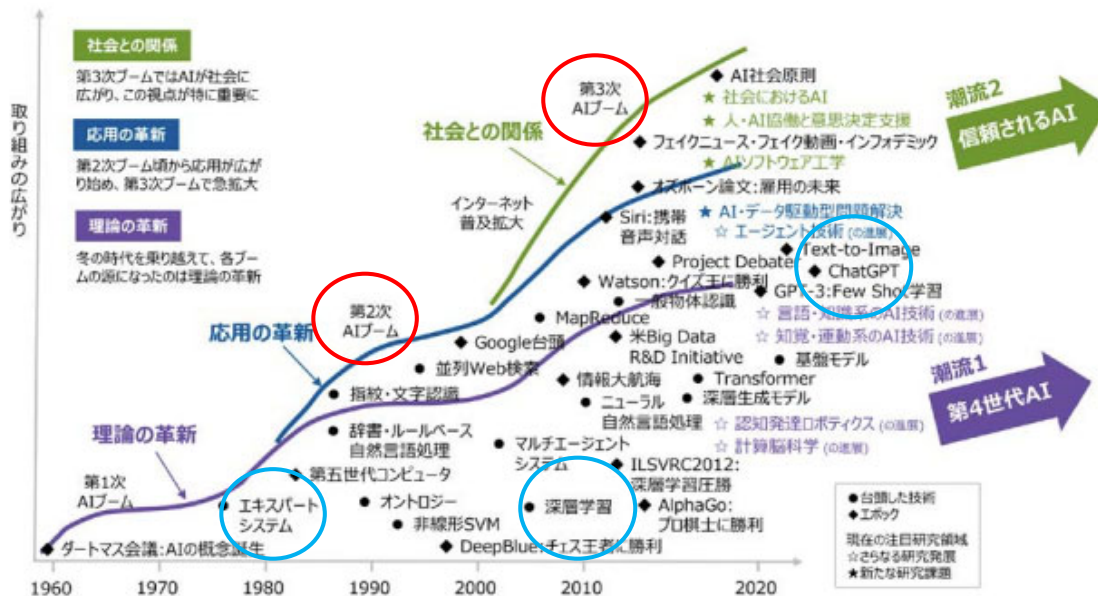
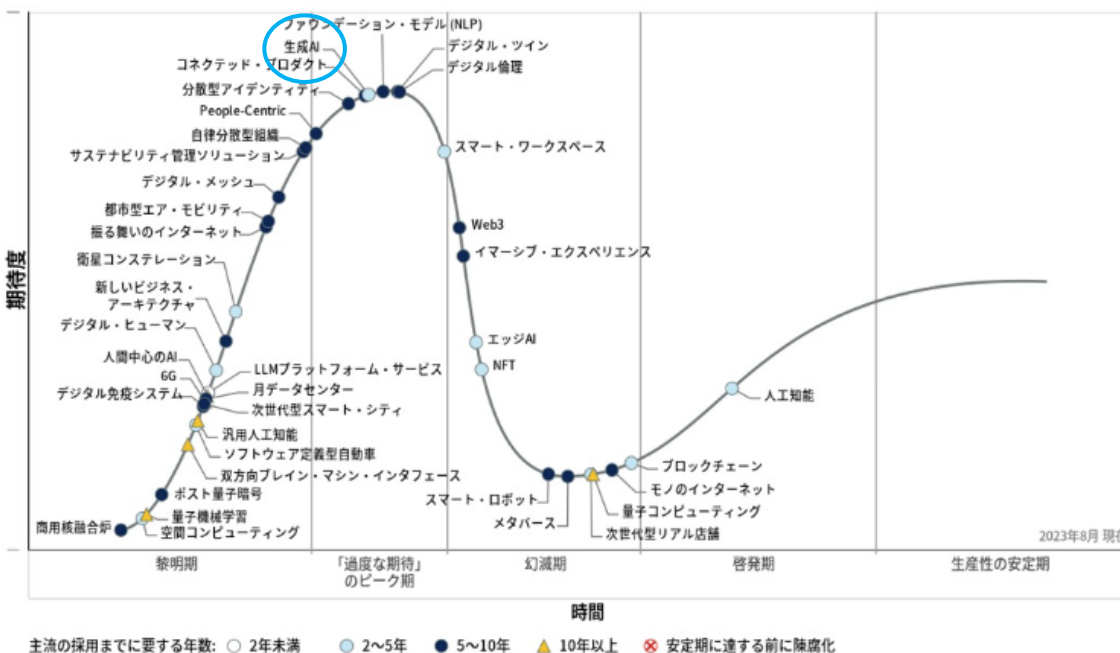


図1-1-1 人工知能・ビッグデータ技術の俯瞰図(時系列)

出典:人工知能研究の新潮流2 ~ 基盤モデル・生成AIのインパクト
国立研究開発法人科学技術振興機構 CRDS

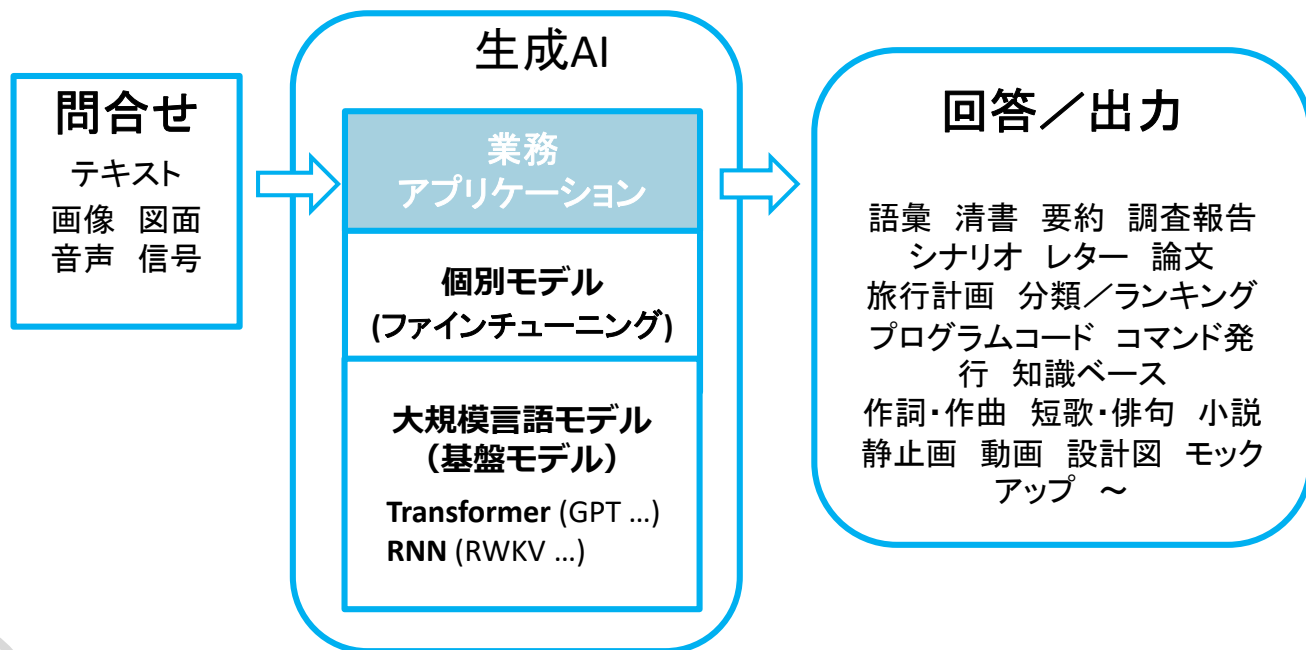
背景: 急速に進展する生成AI

日本における未来志向型インフラ・テクノロジーのハイプ・サイクル: 2023年



出典: Gartner Newsroom 2023年8月17日
<https://www.gartner.co.jp/ja/newsroom/press-releases/pr-20230817>

生成AIの概要



Prompt:

"masterpiece, best quality, ultra-detailed, 8k, high resolution, Draw the landscape painting having the Earth over the pyramid located on the Moon. Earth color is basically blue with white clouds. The space is dark. There are many rocks on the surface of the Moon."

生成AIの概要：導入の課題

生成AIは新しい技術であり、専門家の業務を支援するものとして大きな期待が寄せられていますが、以下に示すような課題があるのも事実です。

適切なLLM(大規模言語モデル)の利用	問題解決プロセスの制御と情報のリスク管理
<ul style="list-style-type: none"> ● ベースとされているデータの内容 ● 基盤モデルのアップデート ● 対象領域に合わせたファインチューニング 	<ul style="list-style-type: none"> ● クラウド利用における内部データの情報漏洩リスク ● 外部リソースの参照に伴う著作権の侵害リスク ● 不適切な出力のチェック

導入支援サービス

STEP1 ヒアリング

STEP2 モデル選定

STEP3 プロトタイプ作成

STEP4 評価／実用化検討

STEP5 実システム開発

STEP6 運用

導入支援サービス:STEP1

- STEP1 **ヒアリング**
- STEP2 モデル選定
- STEP3 プロトタイプ作成
- STEP4 評価／実用化検討
- STEP5 実システム開発
- STEP6 運用

お客様の課題の明確化

- 保有されているデータの簡易分析も実施します



導入支援サービス:STEP2

- STEP1 ヒアリング
- STEP2 **モデル選定**
- STEP3 プロトタイプ作成
- STEP4 評価／実用化検討
- STEP5 実システム開発
- STEP6 運用

課題解決に最適な方式／基盤モデルの選定

- 公開済み基盤モデルの利用を検討します
- 不可の場合は独自の基盤モデル作成を検討します



導入支援サービス:STEP3

- STEP1 ヒアリング
- STEP2 モデル選定
- STEP3 **プロトタイプ作成**
- STEP4 評価／実用化検討
- STEP5 実システム開発
- STEP6 運用

開発環境の構築

- クラウドサービス／オンプレミスの選択をします
- AI開発のライブラリやフレームワークを選定します

ビジネスロジックの組み込み

- 内部／外部のデータアクセスを検討します
- Agent モデルによる自動化を検討します

結果の検証

- お客様と共に入出力内容の妥当性を検証します

導入支援サービス:STEP4

- STEP1 ヒアリング
- STEP2 モデル選定
- STEP3 プロトタイプ作成
- STEP4 **評価／実用化検討**
- STEP5 実システム開発
- STEP6 運用

所定の期間の試用・評価を経た実用化の検討

- クライテリアを設定し、実用化の判断を行います

導入支援サービス:STEP5

STEP1 ヒアリング

STEP2 モデル選定

STEP3 プロトタイプ作成

STEP4 評価／実用化検討

STEP5 実システム開発

実用化目標を定め実用システムの構築

- クラウドサービス／オンプレミスの選択をします

STEP6 運用



導入支援サービス:STEP6

STEP1 ヒアリング

STEP2 モデル選定

STEP3 プロトタイプ作成

STEP4 評価／実用化検討

STEP5 実システム開発

継続的なモデル更新

- データとモデルのバージョン管理を行います
- モデル更新プロセスのパイプライン化を図ります

STEP6 運用



おわりに

生成AIに限らず、広くAIを用いた業務ソリューションの導入・開発から運用まで幅広くお手伝いいたします。

- **当社AIソフトウェアの販売**
 - 第一原理計算ソフトウェア Advance/PHASE
 - ナノ材料解析統合GUI Advance/NanoLabo
 - ニューラルネットワーク分子動力学システム Advance/NeuralMD
 - 深層学習用ツール Advance/iMacle
- **実績**
 1. 深層学習による**時系列データ予測**モデルの作成
 2. 機械学習による**深度推定モデル**実行スクリプトの作成
 3. Advance/iMacleによる中性子回折**スペクトル解析**
 4. ベイズ最適化による自動車1DCAEモデルの**パラメータ同定**
 5. 機械学習による設計条件の最適化
 6. 最適制御のためのオンライン機械学習
 7. PINN : Physics-Informed Neural Networks による**サロゲートモデル**
 8. 学習済みモデルを利用した**webシステム**、**スマホアプリ**の作成

ご清聴いただきありがとうございました。



警告

このレポートに収録されている文章および内容については、ご自身のために役立つ用途に限定して無料配布しています。このレポートを、販売、オークション、その他の目的で利用するには、著作権者の許諾が必要になります。このレポートに含まれている内容を、その一部でも著作権者の許諾なしに、複製、改変、配布を行うことおよびインターネット上で提供する等により、一般へ送ることは法律によって固く禁止されています。