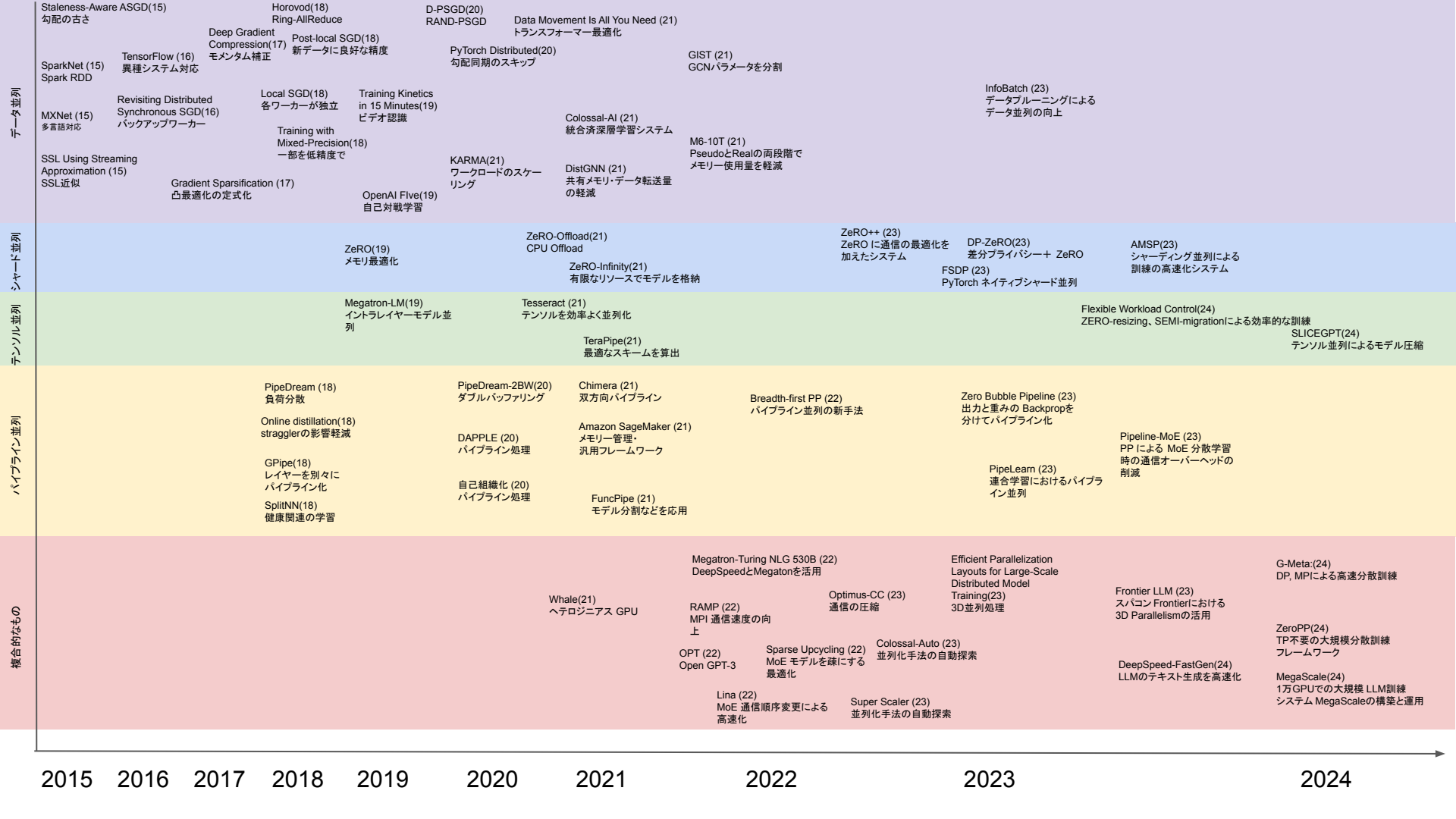
データ並列
シャード並列
テンソル並列
パイプライン並列
複合的なもの
Staleness-Aware ASGD(15)
勾配の古さ
Horovod(18)
Ring-AllReduce
D-PSGD(20)
RAND-PSGD
Data Movement Is All You Need (21)
トランスフォーマー最適化
Deep Gradient
Compression(17)
モメンタム補正
Post-local SGD(18)
新データに良好な精度
TensorFlow (16)
異種システム対応
PyTorch Distributed(20)
勾配同期のスキップ
GIST (21)
GCNパラメータを分割
SparkNet (15)
Spark RDD
Local SGD(18)
各ワーカーが独立
Training Kinetics
in 15 Minutes(19)
ビデオ認識
InfoBatch (23)
データプルーニングによる
データ並列の向上
Revisiting Distributed
Synchronous SGD(16)
バックアップワーカー
MXNet (15)
多言語対応
Training with
Mixed-Precision(18)
一部を低精度で
Colossal-AI (21)
統合済深層学習システム
M6-10T (21)
PseudoとRealの両段階で
メモリー使用量を軽減
SSL Using Streaming
Approximation (15)
SSL近似
KARMA(21)
ワークロードのスケー
リング
DistGNN (21)
共有メモリ・データ転送量
の軽減
Gradient Sparsification (17)
凸最適化の定式化
OpenAI FIve(19)
自己対戦学習
ZeRO-Offload(21)
CPU Offload
ZeRO++ (23)
ZeRO に通信の最適化を
加えたシステム
DP-ZeRO(23)
差分プライバシー＋ ZeRO
AMSP(23)
シャーディング並列による
訓練の高速化システム
ZeRO(19)
メモリ最適化
ZeRO-Infinity(21)
有限なリソースでモデルを格納
FSDP (23)
PyTorch ネイティブシャード並列
Megatron-LM(19)
イントラレイヤーモデル並
列
Tesseract (21)
テンソルを効率よく並列化
Flexible Workload Control(24)
ZERO-resizing、SEMI-migrationによる効率的な訓練
SLICEGPT(24)
テンソル並列によるモデル圧縮
TeraPipe(21)
最適なスキームを算出
PipeDream (18)
負荷分散
PipeDream-2BW(20)
ダブルバッファリング
Chimera (21)
双方向パイプライン
Breadth-first PP (22)
パイプライン並列の新手法
Zero Bubble Pipeline (23)
出力と重みの Backpropを
分けてパイプライン化
Online distillation(18)
stragglerの影響軽減
Amazon SageMaker (21)
メモリー管理・
汎用フレームワーク
DAPPLE (20)
パイプライン処理
Pipeline-MoE (23)
PP による MoE 分散学習
時の通信オーバーヘッドの
削減
GPipe(18)
レイヤーを別々に
パイプライン化
PipeLearn (23)
連合学習におけるパイプラ
イン並列
自己組織化 (20)
パイプライン処理
SplitNN(18)
健康関連の学習
FuncPipe (21)
モデル分割などを応用
Megatron-Turing NLG 530B (22)
DeepSpeedとMegatonを活用
Efficient Parallelization
Layouts for Large-Scale
Distributed Model
Training(23)
3D並列処理
G-Meta:(24)
DP, MPIによる高速分散訓練
Whale(21)
ヘテロジニアス GPU
Optimus-CC (23)
通信の圧縮
Frontier LLM (23)
スパコン Frontierにおける
3D Parallelismの活用
RAMP (22)
MPI 通信速度の向
上
ZeroPP(24)
TP不要の大規模分散訓練
フレームワーク
Colossal-Auto (23)
並列化手法の自動探索
OPT (22)
Open GPT-3
Sparse Upcycling (22)
MoE モデルを疎にする
最適化
DeepSpeed-FastGen(24)
LLMのテキスト生成を高速化
MegaScale(24)
1万GPUでの大規模 LLM訓練
システム MegaScaleの構築と運用
Lina (22)
MoE 通信順序変更による
高速化
Super Scaler (23)
並列化手法の自動探索
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024