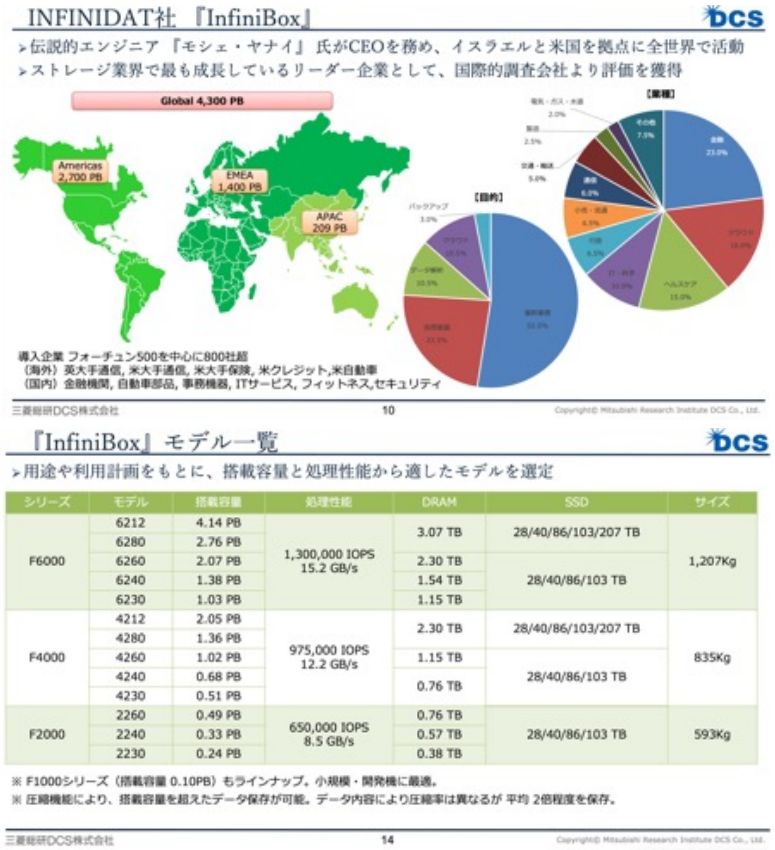
電源10KVA　空調6KW　欲しい

金融系などハイエンドストレージで
大きなシュア（半分？）を持つ『InfiniBox』



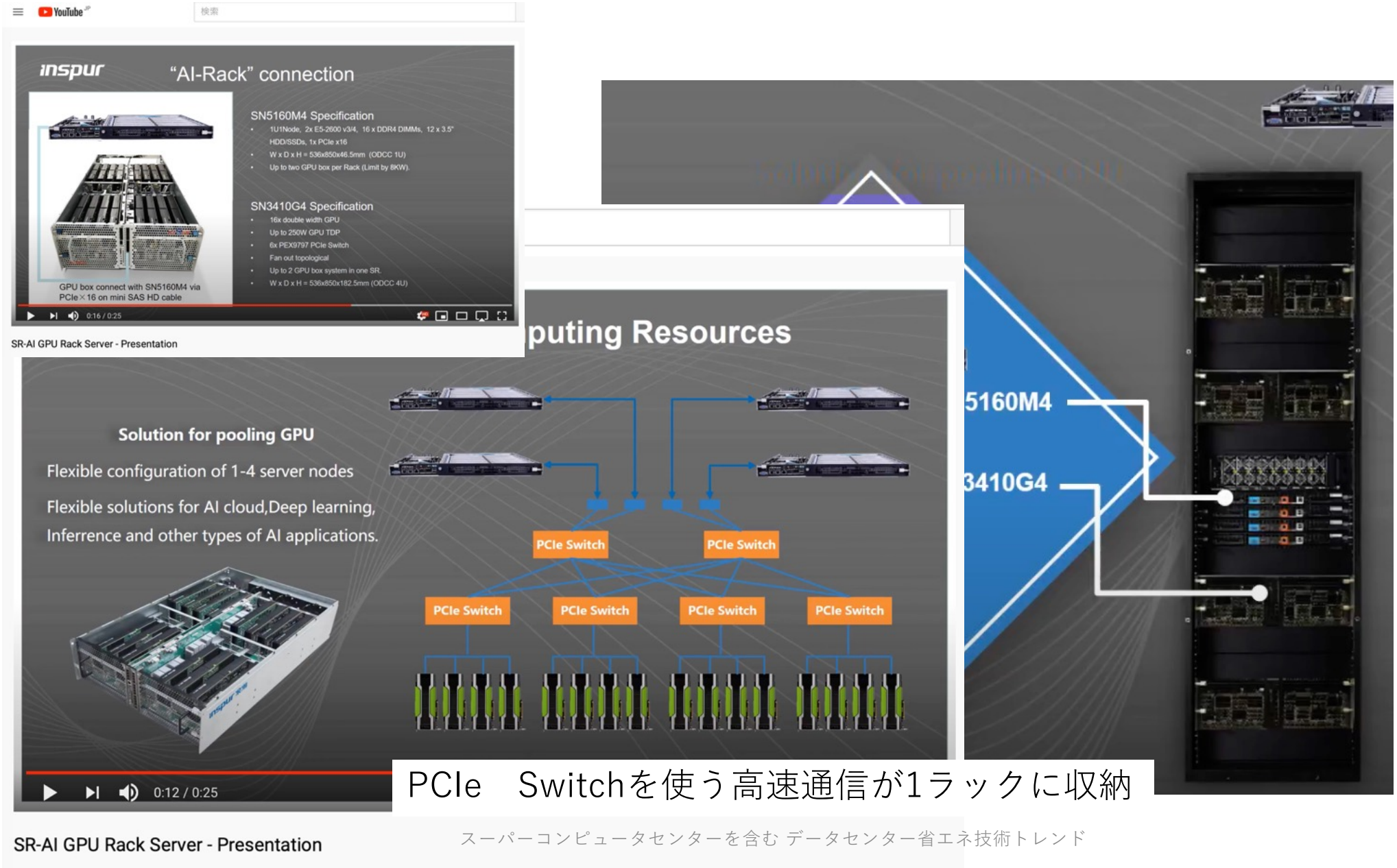
シリーズ	モデル	搭載容量	処理性能	DRAM	SSD	サイズ
F6000	6212	4.14 PB	1,300,000 IOPS 15.2 GB/s	3.07 TB	28/40/86/103/207 TB	1,207Kg
	6280	2.76 PB				
	6260	2.07 PB		2.30 TB	28/40/86/103 TB	
	6240	1.38 PB		1.54 TB		
	6230	1.03 PB		1.15 TB		
F4000	4212	2.05 PB	975,000 IOPS 12.2 GB/s	2.30 TB	28/40/86/103/207 TB	835Kg
	4280	1.36 PB				
	4260	1.02 PB		1.15 TB	28/40/86/103 TB	
	4240	0.68 PB		0.76 TB		
	4230	0.51 PB				
F2000	2260	0.49 PB	650,000 IOPS 8.5 GB/s	0.76 TB	28/40/86/103 TB	593Kg
	2240	0.33 PB		0.57 TB		
	2230	0.24 PB		0.38 TB		

最上位シリーズは1ラック12KVA
引用先：**GUTP/JDCC 「次世代データセンター勉強会」**

稼働率を60%としても　電源10KVA
冷却能力は　1ラック6KWが欲しい

ここまでは、過去のこと

世界で一番たくさんGPUを売っている

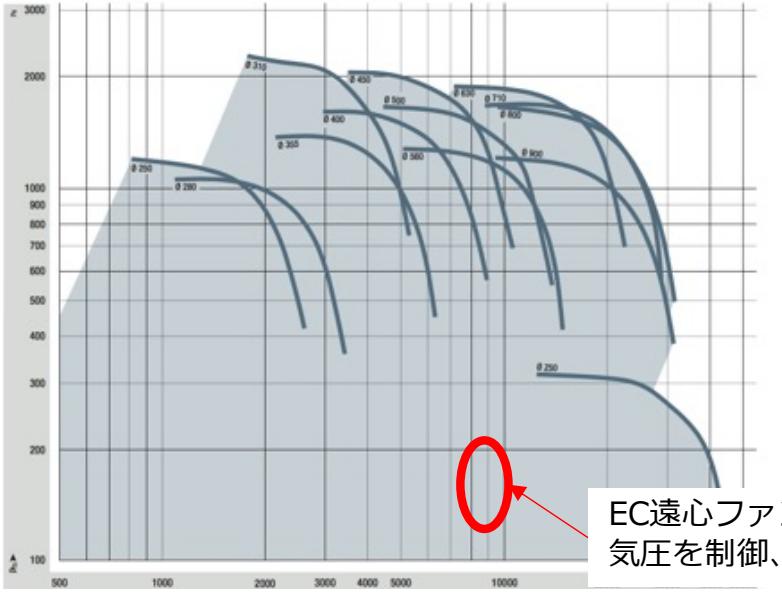


PCIe　Switchを使う高速通信が1ラックに収納

東大「次世代データセンター勉強会」30KVAラックの設計ガイドライン

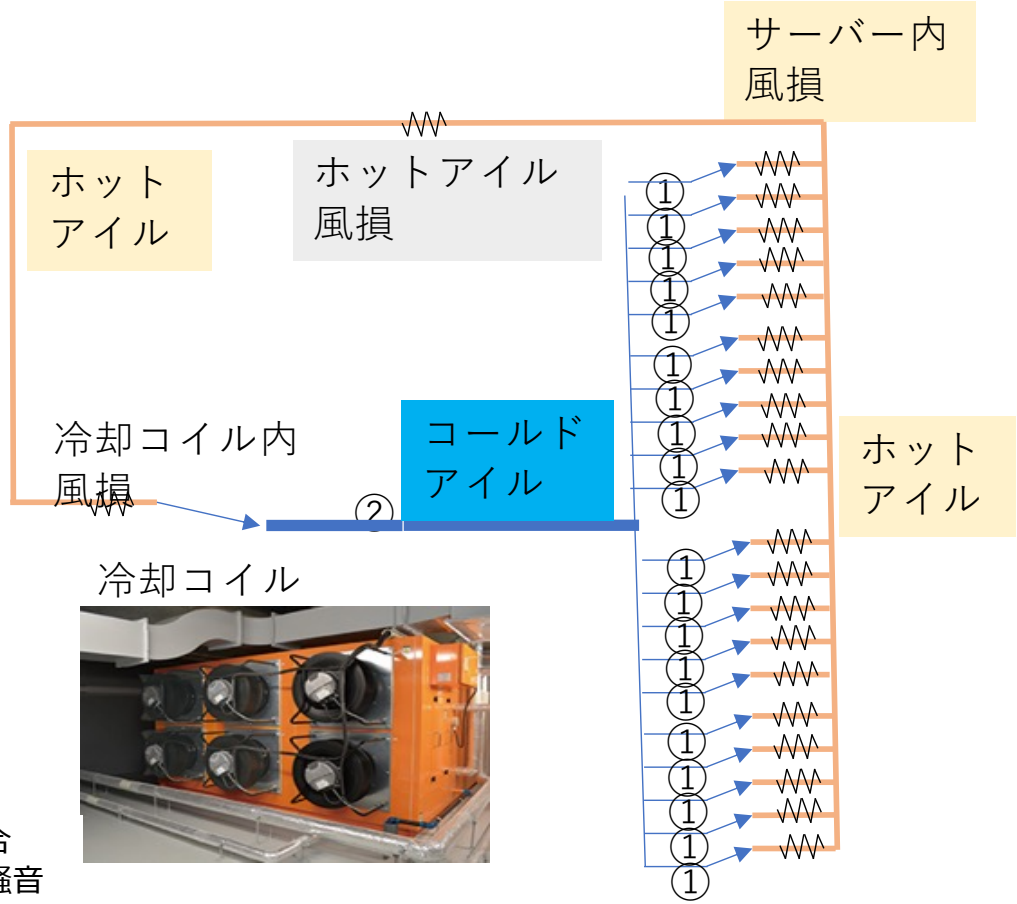
必要な風量は消費電力に比例する

10KW20ラック　34,000m3/hを6台で供給とすると
おおよそ@6,000m3/hを送風可能なファンを選択
ただし風損で風量が不足しないこと。



EC遠心ファンΦ400機が適合
気圧を制御、低速稼働し低騒音

引用先：http://www.ebmpapst.jp/pdf/RadiPac_US_18_12_201502_J.pdf

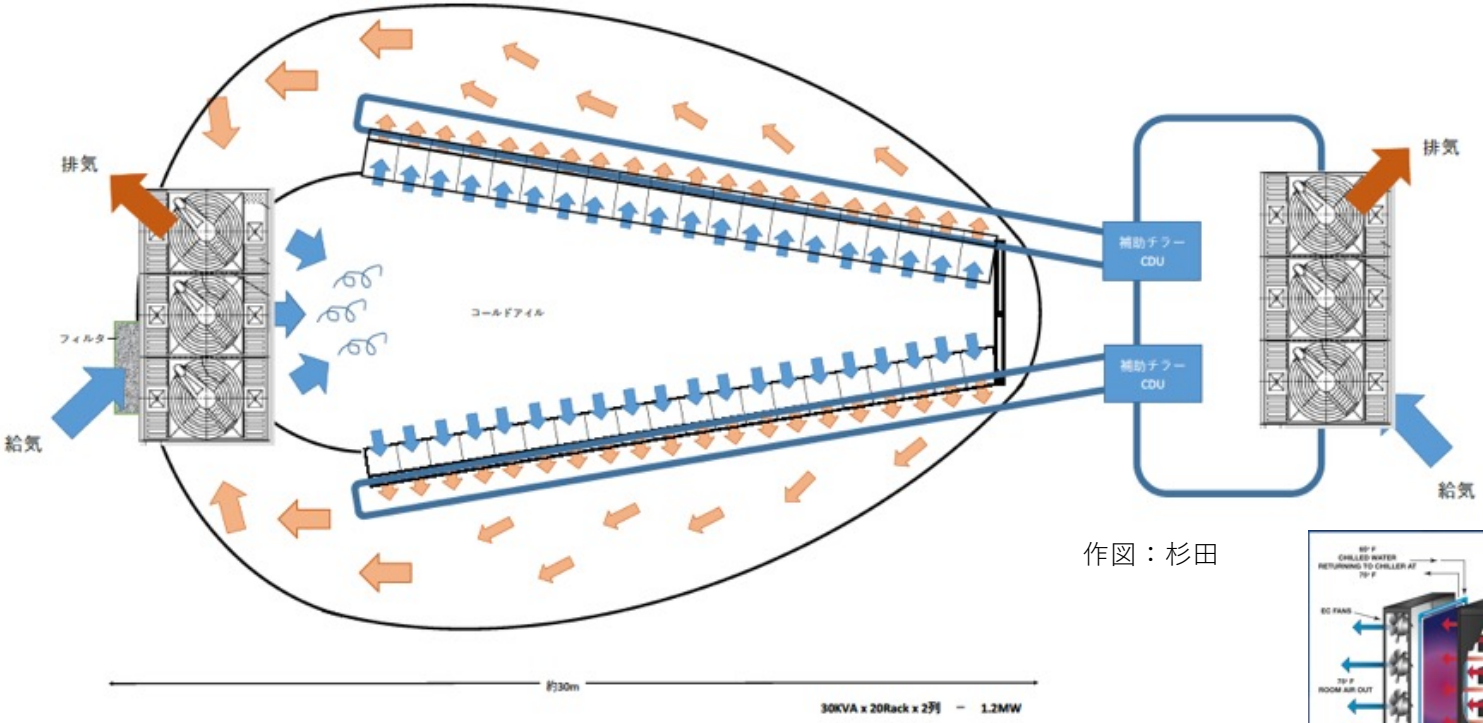


引用先：https://www.chiroro.co.jp/datacenter/

30KVA/1Rack　1.5MWDCデザイン
40Rack 間接外気冷却システムと
リアドアとの2重ループ冷却

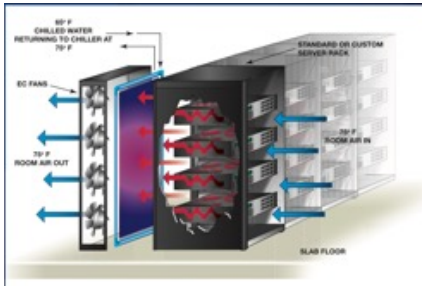
重要

④ サーバ視点から見た熱力学の法則遵守



作図：杉田

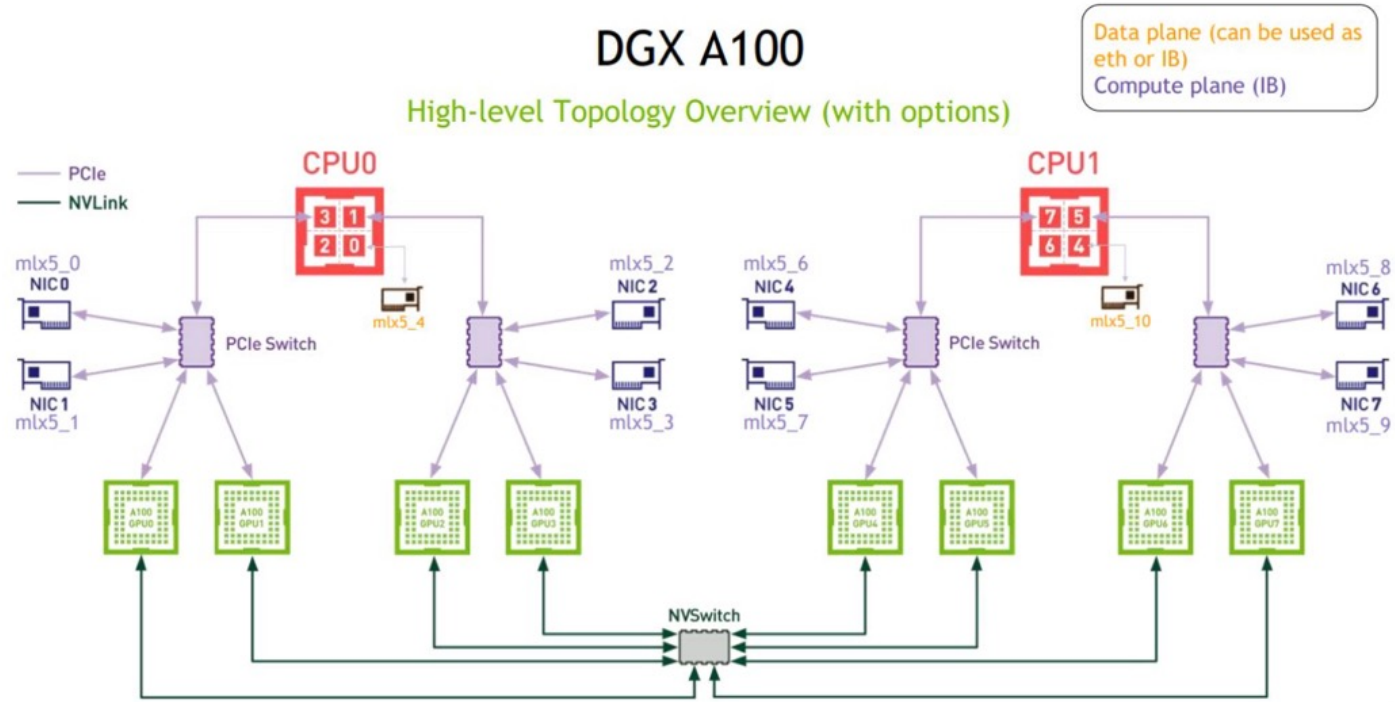
内蔵ファン静圧が低い１U機を禁止して、リアドアとの冷却2重ループとして「ラック通過風速」を低く出来る。



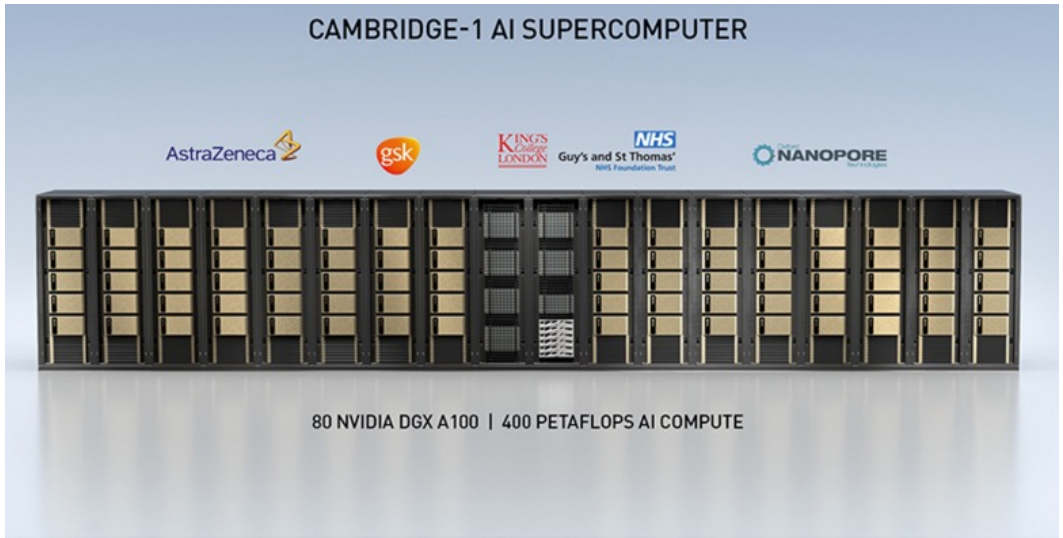
リアドア

風量と風速（1m/s）、気圧を考慮した　たまご型
電気容量1.5MW　空調容量1.2MW

引用先：
http://www.chilleddoor.com/files/uploads/2016/10/MOT_chilledDoorHR_11-2016_REVnoflap.pdf

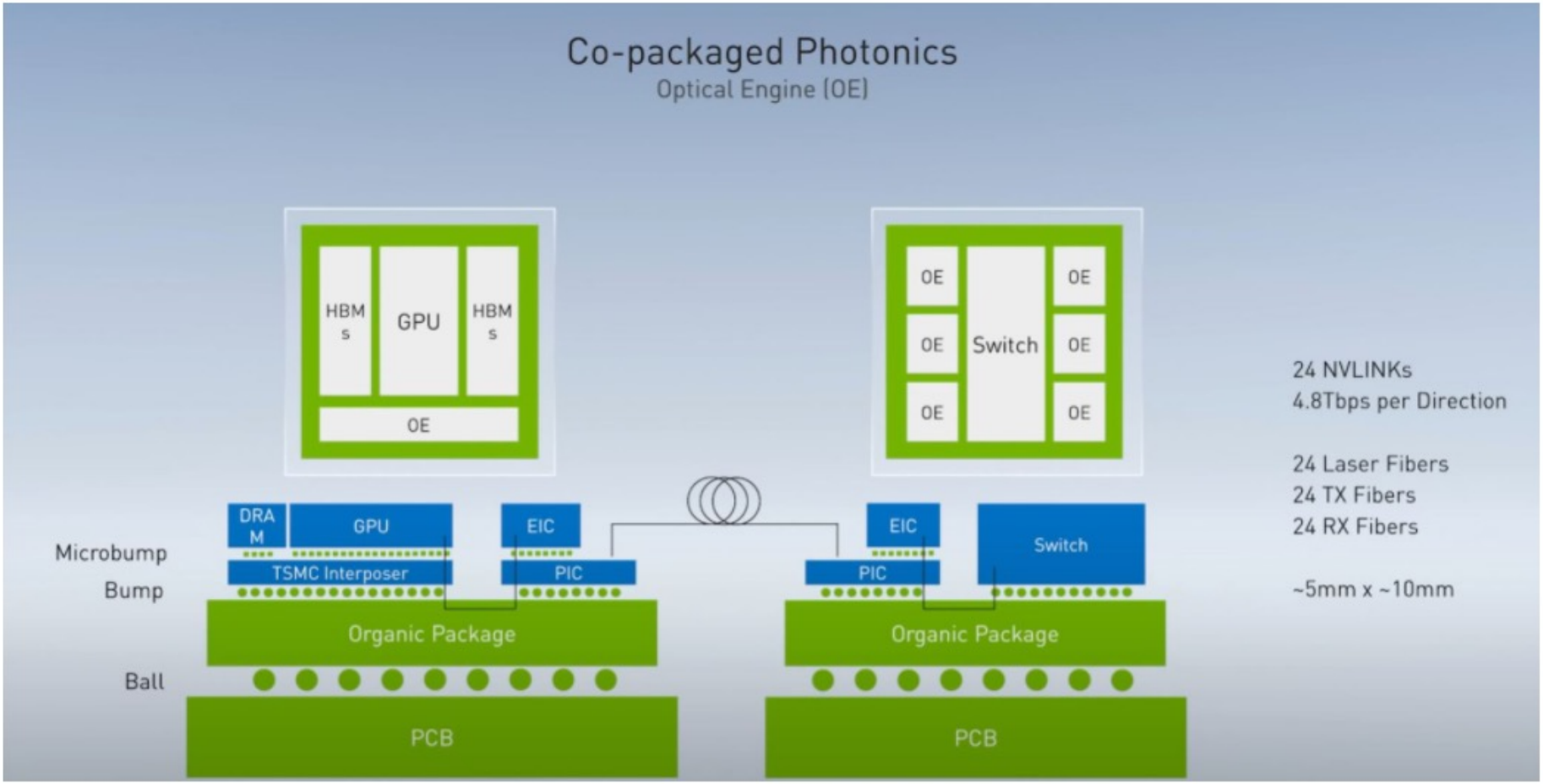


HC32 NVIDIA DGX A100 SuperPOD DGX A100 System Topology



普通のデータセンターでは
排熱が難しい。
左の構成は稼働するが、
回りに何も設置されていない。

チップとチップを光ファイバで結ぶ



引用先：https://www.servethehome.com/nvidia-gtc-china-2020-co-packaged-photonics-and-inference/?fbclid=IwAR1oLhTZlOGxc6YEfEnsJX0X6iaNawpXWUowgES8TNl3NhrFIOHGkkcpKrI

AI & 機械学習

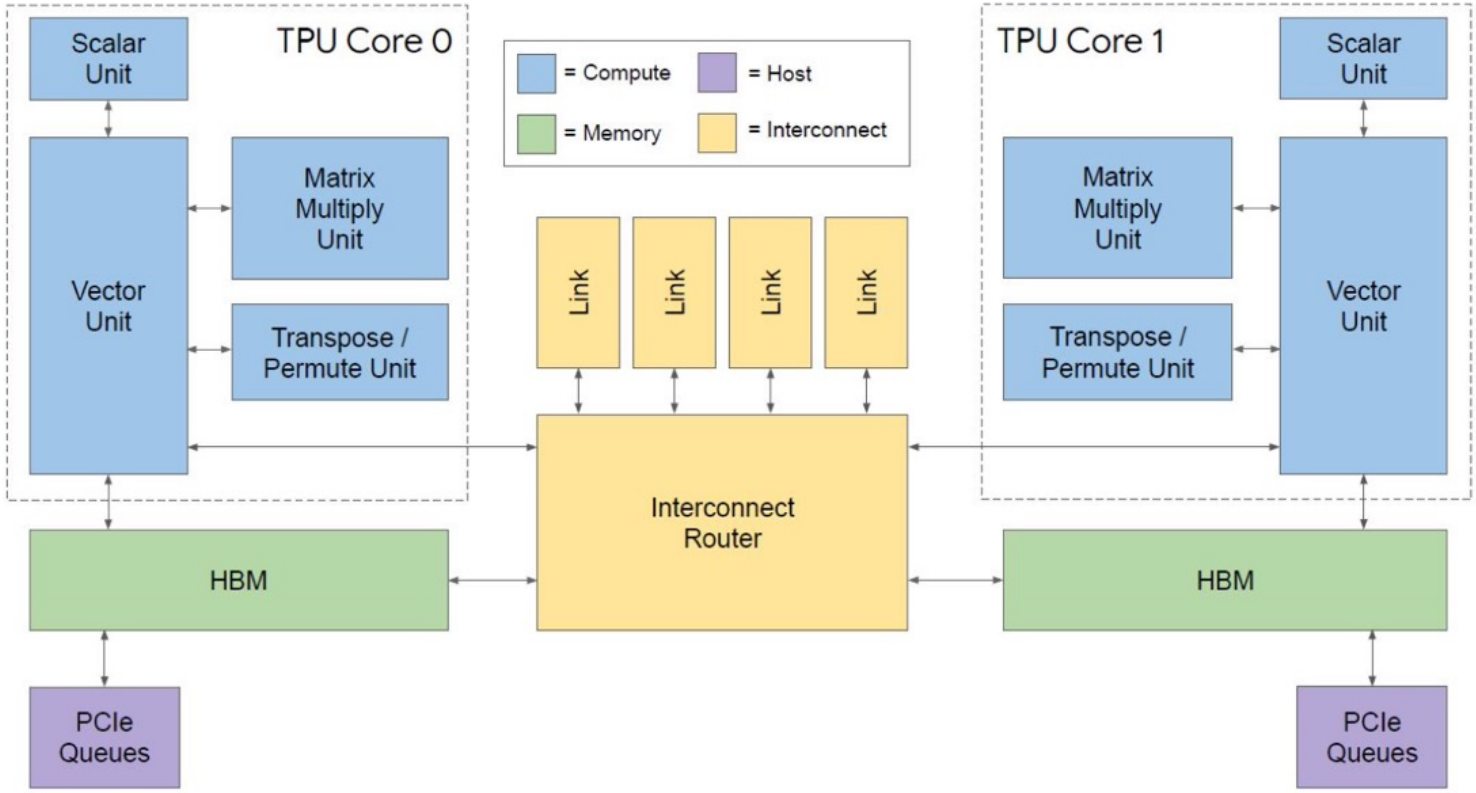
Google が世界最速のトレーニング スーパーコンピュータで MLPerf の AI パフォーマンスの記録を塗り替える

Google Cloud Japan Team

※この投稿は米国時間 2020 年 7 月 30 日に、Google Cloud blog に投稿されたものの抄訳です。

TPUv3

https://www.servethehome.com/google-tpuv3-discussed-at-hot-chips-32/?fbclid=IwAR0RiFopRO9LtkUmGOOHByj5DJmXsWft9hP4XmbxUiuKno9k7nQ8N8xluH0

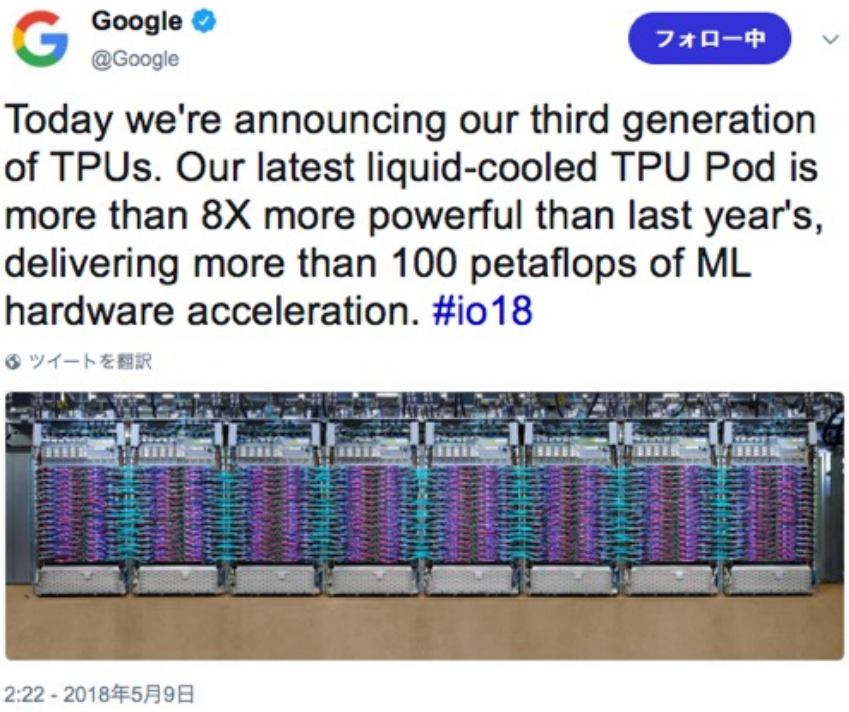


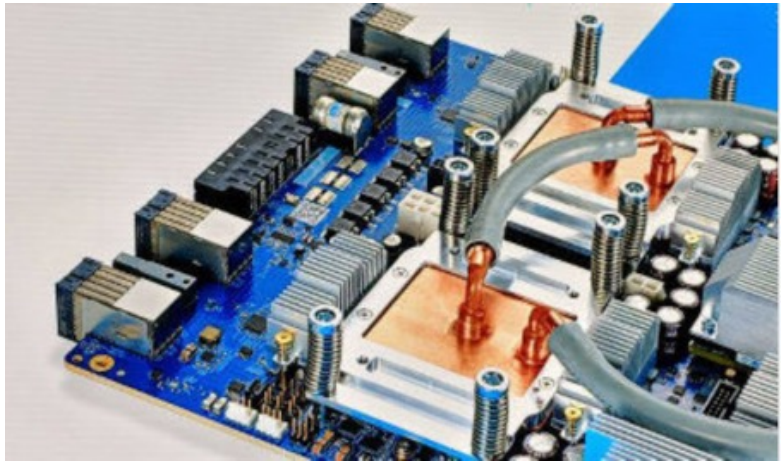
HC32 Google TPUv3 Overview Diagram

本年TPUv4を使うスパコンが発表された

Googleのスパコン　TPU3.0

超高速ネットワーク
詳細は未公表



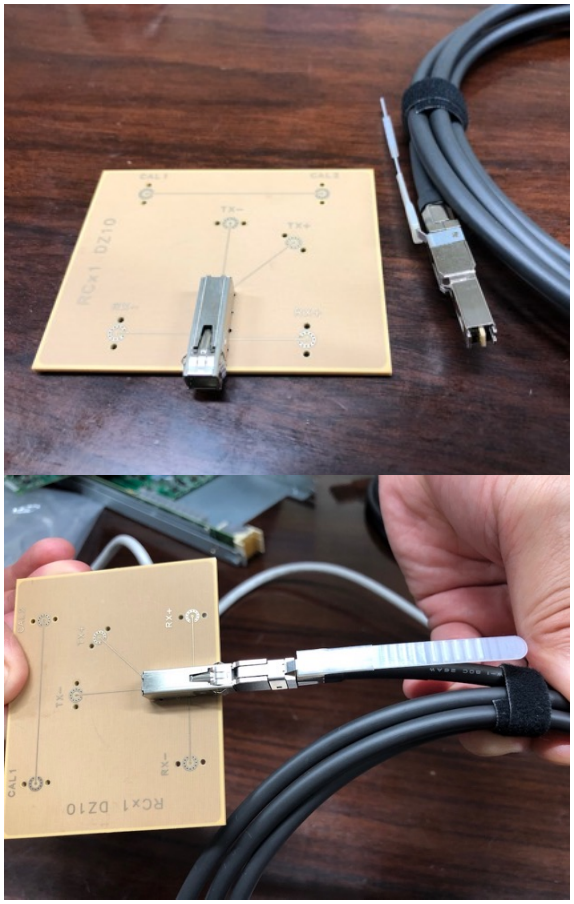


引用先：https://gigazine.net/news/20180510-google-tpu-3-ai-processor/

第3世代のTPU3.0では、さらに「トレーニング」に要する時間を大幅に短縮できる

TPU2.0と比べ8倍密度、たった8ラックで100ペタフロップス

データセンターはサーバーを置く場所



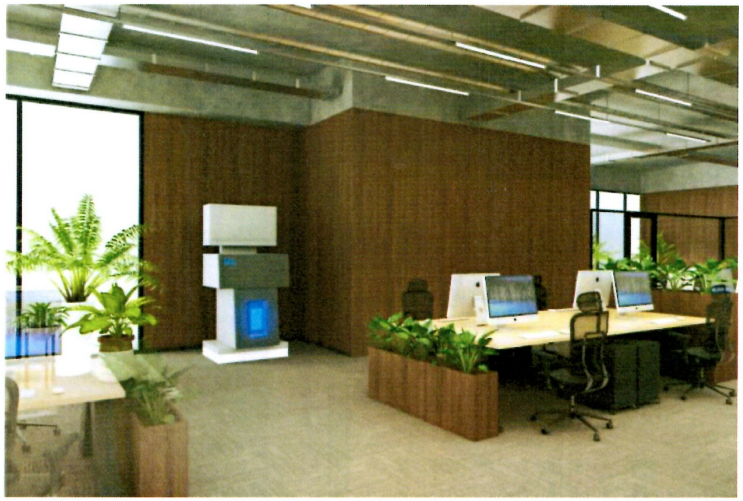
25Gbpsカッパーコネクタ

25Gbpsカッパー　ケーブル

※日本人営業がマイクロソフトに、このコネクタを売り込みに行きました。
マイクロソフトは3ヶ月で25万台サーバーを買うとのこと。
Goolge、AWSはMSより買っているとのこと
・・・日本市場は全体で3ヶ月10万台規模・・・

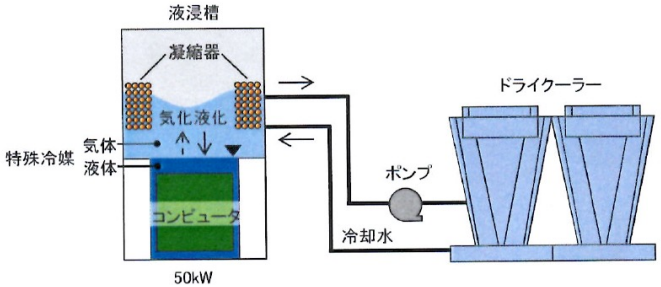
オフィスに設置出来るスパコン「液浸」

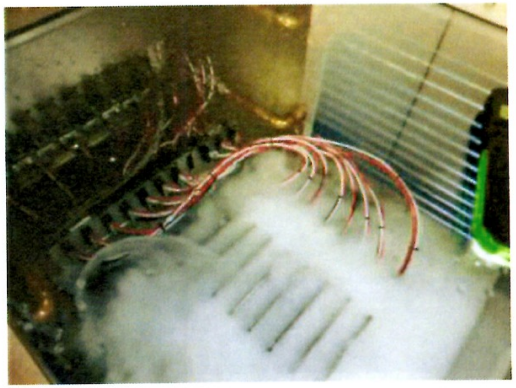
オフィス設置イメージ



ビル設置イメージ

システム構成図





引用先：大成建設

近い未来

- Future

半導体は進化して　さらに電気を食べる

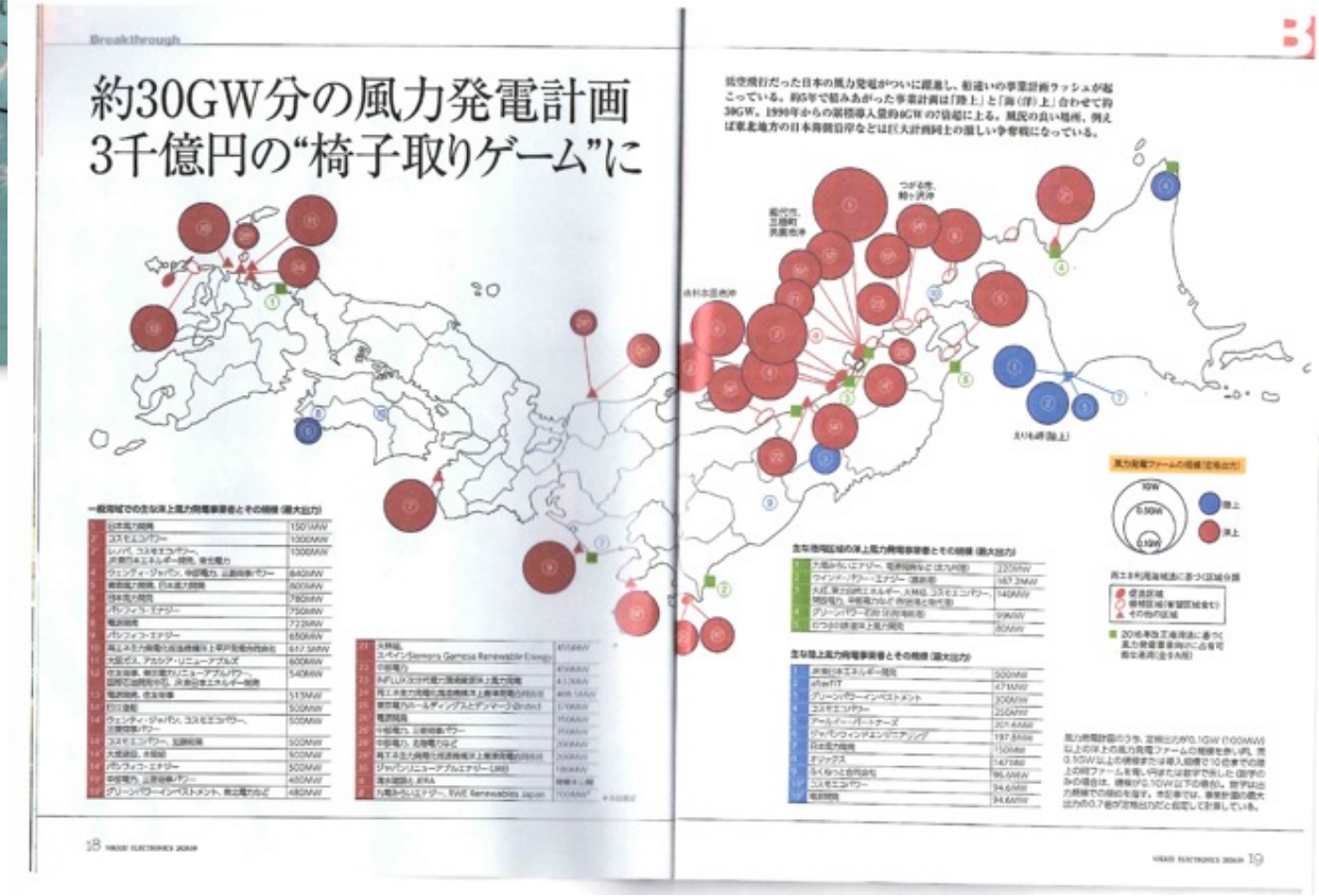
NVIDIA　V100は　台湾TSMCが製造できる最大サイズのチップ　TDP=300W	
PFNが作ったMN3は　TDP=600W	
ウエハースケール（ウエハーを切断しない）は、TDP=20KW　20,000KW	
	https://cn.teldevice.co.jp/product/cerebras_cs-1/

みんな忘れている「サーバーはデータセンターに置く」ことを。

データセンターこそ
ゼロエミッション
再生可能エネルギーによる
データセンター運用には
先ず　省エネ　が必須です。

日経エレクトロニクス　2020年9月号
特集　「いきなり風力発電大国」

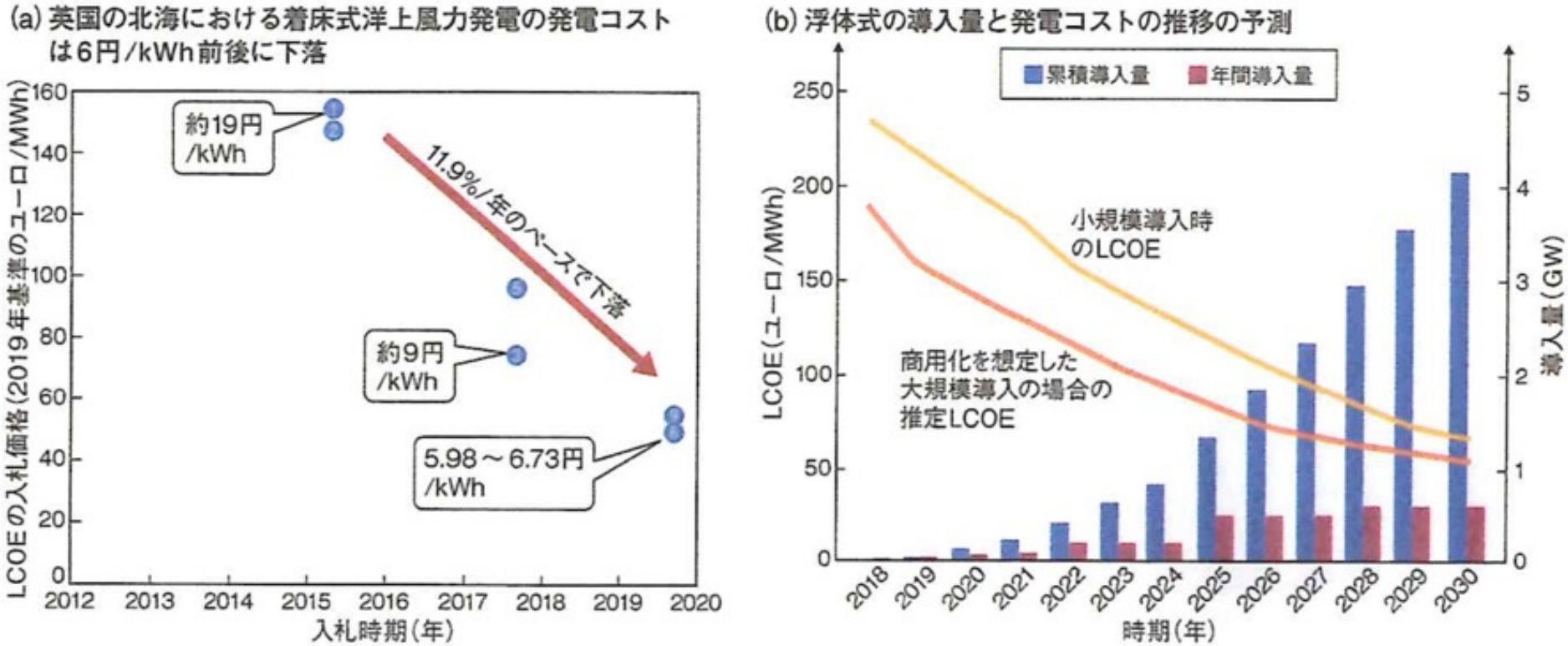
洋上風力発電が実用化
1KWh7円以下を実現へ



当然　安価な
大規模蓄電池も必須

日経エレクトロニクス　2020年9月号
特集　「いきなり風力発電大国」

Breakthrough



LCOE：Levelized Cost of Energy

図4　洋上風力の発電コストが急降下

北海の風力発電ファームにおける着床式(a)と浮体式(b)の洋上風力発電の発電コスト(LCOE)の推移や予測を示した。着床式は2015年ごろは約19円/kWhと高かったが、そこから年間11.9%減のペースで下落し、2019年9月の入札では51ユーロ/MWh(6.3円/kWh)と非常に安い水準になった。浮体式のLCOEは、現状では着床式の2～3倍と高いが、導入量の増加に応じて急速に下落し、2023～2024年には大規模導入を想定すると火力発電並みのコスト、2030年ごろには現在の着床式と同等の水準になる見通しだ。(図：英Imperial College London の研究者の論文[1]のデータを基に日経エレクトロニクスが加筆、(b) WindEurope, "Our energy, our future", Nov. 2019より引用)

さくらインターネット
石狩データセンター　直流給電システム

この直流太陽光システムは、高い電圧から低い電圧に単純に電流が流れるだけのシンプルな仕組みであります。交流電力のような切り替え制御をしなくても無制御で構成できています。

例えば、晴れているときは太陽光の電圧が一番高いため、太陽光から電流が流れる。天候が悪くなり暗くなれば太陽光の電圧は下がるため、次に高い電圧は系統電力となりそれを直流に変換した電圧から電流が流れる事になります。

系統が停電した場合、次に高い電圧はバッテリー電圧のためバッテリーから電流が流れることになります。（図 2.5-1）。

直流は、高い電圧から低い電圧に川の流れのように電流が流れます。本システムはこれらの直流の特性を最大限に生かしているため、シンプルな構成であり、無制御で高効率・高信頼性のエネルギー供給システムを実現しています。

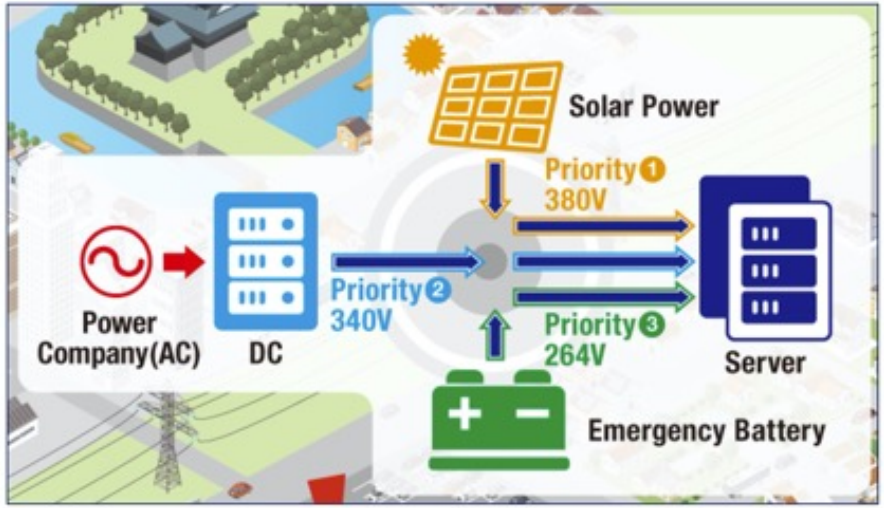
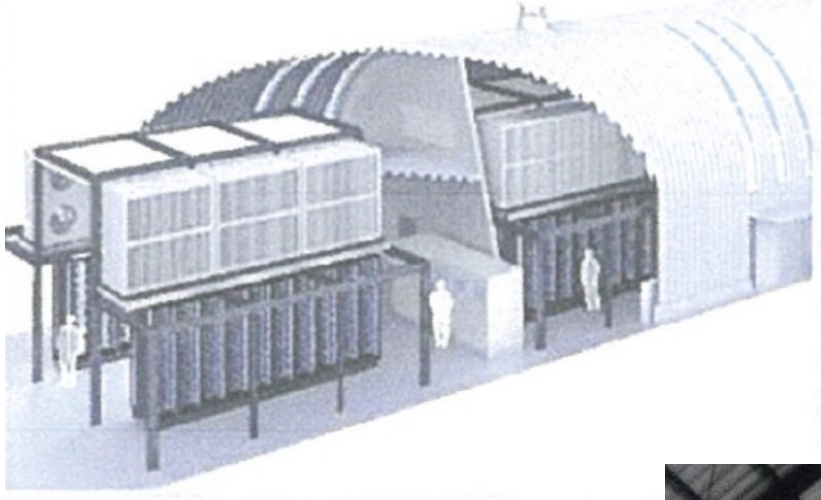


図 2.5-1　単純な直流接続で高い電圧から低い電圧に流れる無制御システム

マイクログリッドとして
他にもNEDOで実証試験が
行われている。

出所：NTTデータ先端技術株式会社

建物コストを最小に、PUEも低い



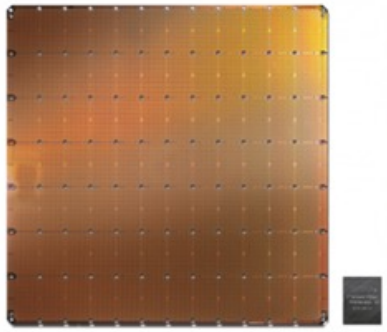
未来

- Future

  半導体は進化して
- さらに電気を食べる

大型半導体　～ウェハースケールエンジン(WSE)～

「CS-1」には、ディープラーニング専用に設計された21.5cm角の大型半導体「WSE(Wafer-Scale Engine)」を搭載しています。「WSE」は積和演算を実行する40万個のコアと、効率的なデータアクセスのためにコア毎に高速なローカルメモリを実装し、コア間で非常に高速な通信が可能になるように設計されています。



46,225㎟ チップ
最大のGPUの56倍の大きさ

400,000コア
最大のGPUの78倍以上

18 GB オンチップSRAM
最大のGPUの3,000倍以上

100 Pb/s コア間接続
最大のGPUの33,000倍以上

ディープラーニング専用に設計された21.5cm角の大型半導体「WSE(Wafer-Scale Engine)」

ディープラーニングに最適化

WSEの40万個の疎線形代数演算コアは、多くのニューラルネットワーク向けに最適化されています。各コアはそれぞれ独立して柔軟にプログラム可能であるため、トレンドに合わせてどんなアルゴリズムでも実行することができます。

広帯域・低遅延な通信ファブリック

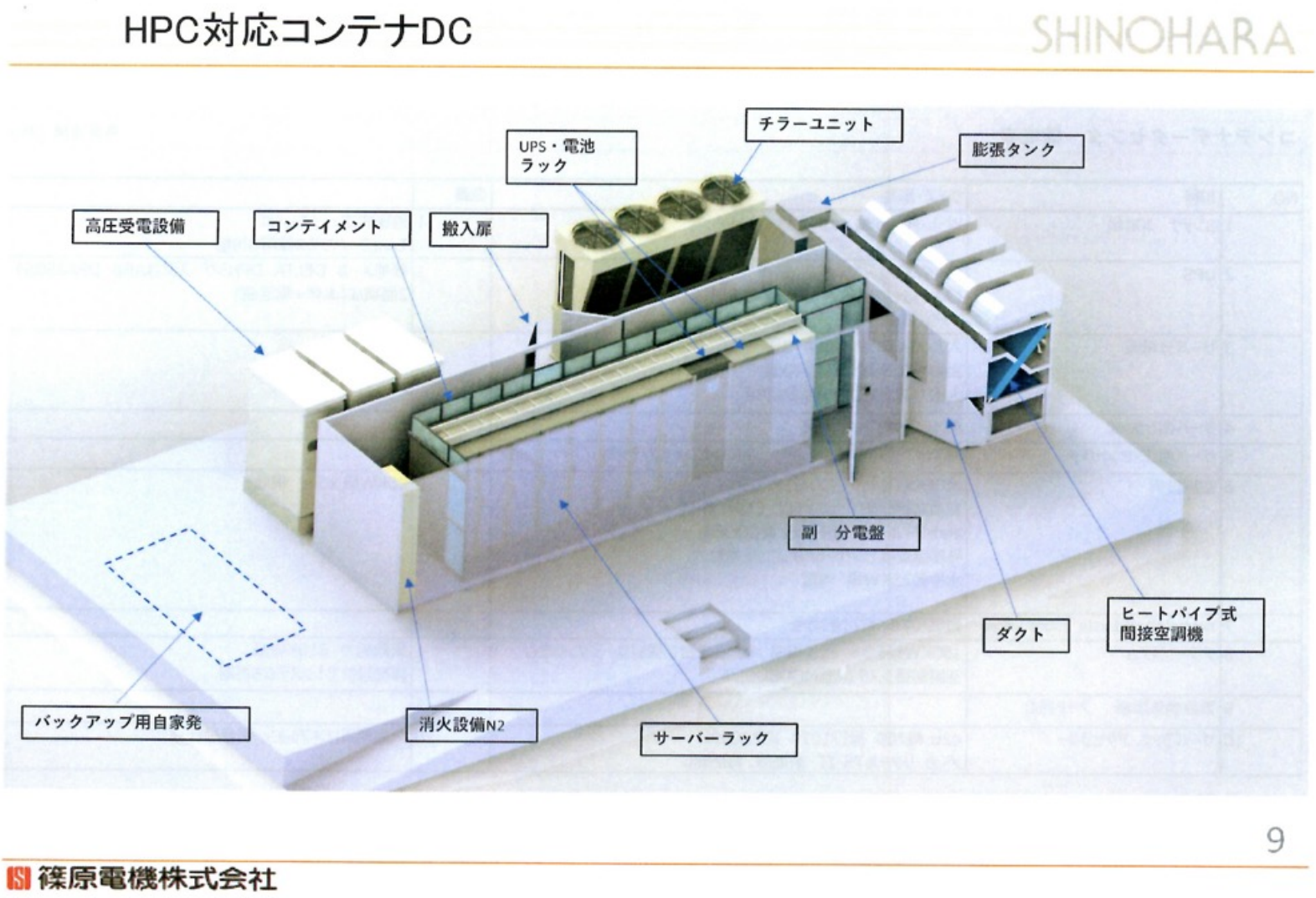
WSE上のコアは、Swarmと呼ばれる100Pb/sの2次元メッシュ構造の100Pb/sの通信ファブリックで相互に接続されています。SwarmによりGPU同士間の通信と比べて非常に小さなレイテンシと消費電力で広帯域通信することが可能です。また、ユーザー仕様のモデルを学習させるためにWSE上でのコア間の最適な通信経路を設定します。

高速なオンチップメモリの使用

WSEは合計18GBのオンチップメモリを実装しており、9.6PB/sのメモリバンド幅でアクセスします。これは、業界を代表する他社製品と比べて3,000倍以上のメモリ量、10,000倍以上のメモリバンド幅があることを意味します。より多くのコアにより多くのローカルメモリを配置することで、より低レイテンシで低消費電力を実現しながら高速で柔軟な演算処理が可能になります。

みんな忘れている「サーバーはデータセンターに置く」ことを。

内部にHPC対応コンテナ



データセンターの再エネ利用は進めないといけない。

これから10年でデータセンター世界需要は2倍から
3倍になると予測されている。
全世界の電力需要の３％から５％に達する。
これを石炭・石油などによる発電でまかなうのは無理がある。
太陽光発電、洋上風力発電が安価になり、充電設備コストが
下がると、石炭・石油による発電を利用するデータセンターは
駆逐されてゆくだろう。

データセンターの再エネ利用は進めないといけない。

GAFAと言われるクラウド大手は、サービス性を上げるため益々データセンター設備を増強するだろう。高処理のための大規模データセンター利用は他社も追従するだろう。
しかし、クラウド大手は、会社存亡リスクとして、再生可能エネルギーに投資している。現状は主に債権スワップで再生可能エネルギー100%を実現したと言っている。

データセンターの再エネ利用は進めないといけない。

東京は、わずか450ほどのビルが
東京都の15%も電気を使っている。
日本の総電力の2%以上　データセンターが使っている。

中国でもITが使う電力量は問題になっていて、豪雨で決壊が心配される三峡ダムが発電する総電力量をITが使っていると問題化している。
北京市内には新たなデータセンター建設が許可されていない。

以上