

令和3年度重要技術管理体制強化事業
(マイクロエレクトロニクスに係る
産業基盤実態等調査)
調査報告書

令和4年3月31日

NTTアドバンステクノロジー株式会社

目次

はじめに	2
1. マイクロエレクトロニクスに係る市場動向	5
1-1. 世界電子機器市場および半導体関連市場概要の状況	6
1-2. 主要なマイクロエレクトロニクスに係る製造メーカーおよび製造装置メーカー、材料メーカーに係る市場動向、主要用途分析、技術動向、生産開発拠点及び生産能力構築状況	13
1-3. 主要生産国におけるマイクロエレクトロニクスに係る生産に対する公的支援の実績と効果の調査	198
2. マイクロエレクトロニクスに係る技術動向	248
2-1. 世界のマイクロエレクトロニクスに係る開発・応用技術に係る先端的技術の開発動向の調査	249

はじめに

本調査では、マイクロエレクトロニクス分野の技術動向並びに競争環境に関する情報に基づいて、内外の研究開発動向を踏まえつつ、それら技術の研究基盤・産業基盤の実態等を把握することを目的とする。

企業名の表記について

本報告書では、文脈に応じて英語・日本語・その他現地語などの言語や原文の表記を尊重する場合など、スライドによって企業名の表記が異なる場合があります。下記にリストアップするので、参考にされたい。

代表的表記	日本語表記	英語表記	中国語表記
インテル		Intel	
TSMC		Taiwan Semiconductor Manufacturing Company	台湾積体電路製造
サムスン	サムスン電子、三星、三星電子	Samsung、Samsung Electronics	
SK ハイニックス	ハイニックス	SK Hynix、Hynix	
SMIC		Semiconductor Manufacturing International Corporation	中芯国際
YMTC	長江ストレージ	Yangtze Memory Technologies Corp、Yangtze River Storage Technology	长江存储科技
華虹グループ		Huahong Group	(上海)華虹集團
紫光集團		Tsinghua Unigroup	清華紫光集團
インフィニオン		Infineon	

1. マイクロエレクトロニクスに係る市場動向

1-1. 世界電子機器市場および半導体関連市場概要の状況	6
1-1-1. 調査の目的と背景	7
1-1-2. 世界の電子情報機器市場動向	9
1-1-3. 世界半導体市場売上予測	11
1-2. 主要なマイクロエレクトロニクスに係る製造メーカーおよび製造装置メーカー、材料メーカーに係る市場動向、主要用途分析、技術動向、生産開発拠点及び生産能力構築状況	13
1-2-1. 調査の目的と背景	14
1-2-2. Foundry市場	18
1-2-3. IDM市場	42
1-2-4. Fabless市場	73
1-2-5. OSAT市場	81
1-2-6. EMS市場	106
1-2-7. 半導体製造装置市場	120
1-2-8. 半導体材料市場	150
1-2-9. まとめ	194

1-1.世界電子機器市場および半導体 関連市場概要の状況

1-1-1. 調査の目的と背景

【現状】

- COVID-19の影響による半導体製造工場の一時的な閉鎖や、米国の中国への先端半導体技術の規制に端を発する半導体サプライチェーンのデカップリング、また、COVID-19に起因する電子機器の需要増などさまざまな要因により半導体不足が顕著になった。

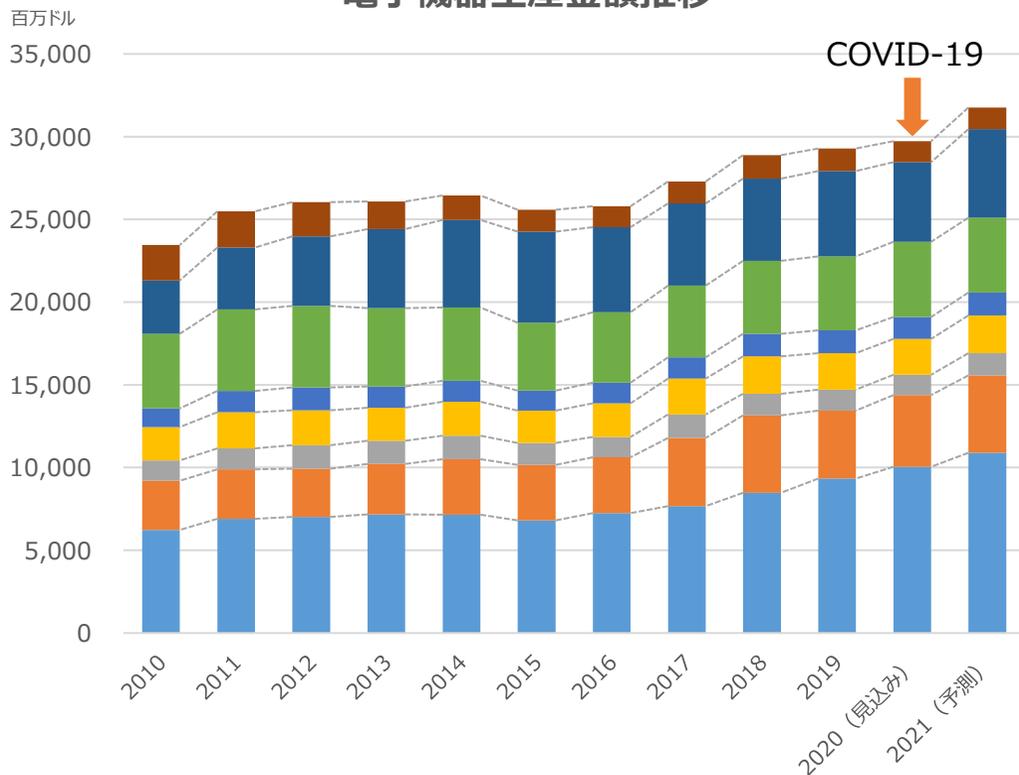
【目的】

- マイクロエレクトロニクスの最終製品である電子機器市場及び半導体市場を定点観測することで、マイクロエレクトロニクス関連市場の基礎的な動向を把握する。

1-1-2. 世界の電子情報機器市場動向

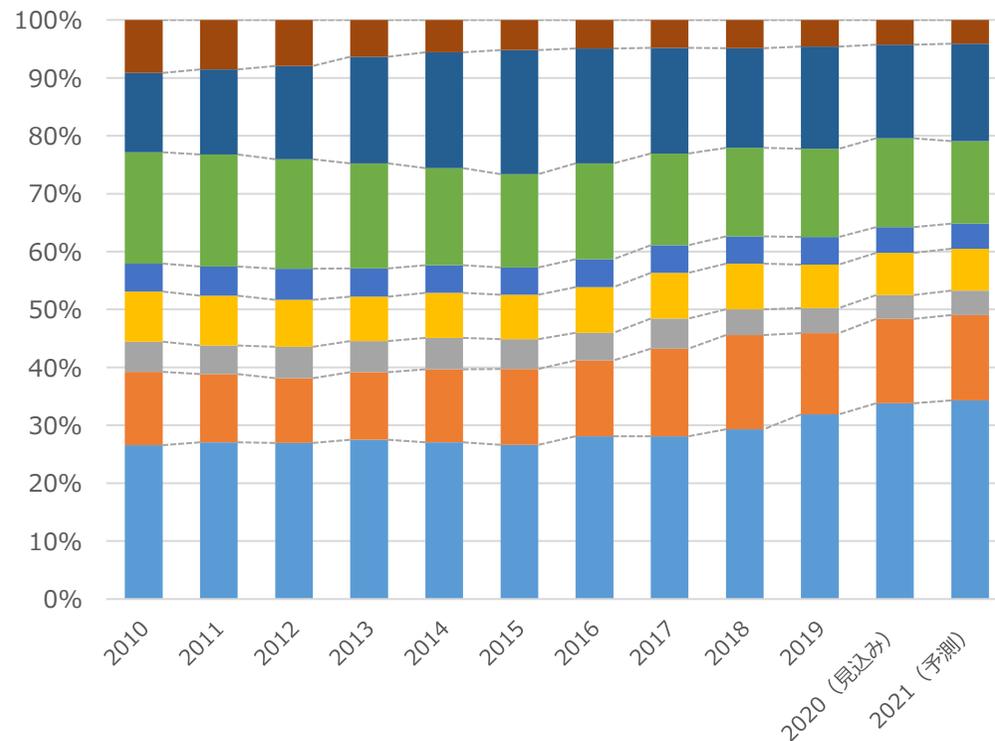
- COVID-19の影響で2019年から2020年にかけてソリューションサービスが伸びていることにより電子機器全体で微増となっている。また、2021年には大きく回復すると見られている。
- 半導体はCOVID-19に大きく影響されず、生産金額が伸びている。

電子機器生産金額推移



- ソリューションサービス
- 電子部品
- 通信機器
- 半導体
- その他電子機器
- AV機器
- ディスプレイデバイス
- コンピュータ及び情報端末

電子機器生産金額推移(割合)



- ソリューションサービス
- 電子部品
- 通信機器
- 半導体
- その他電子機器
- AV機器
- ディスプレイデバイス
- コンピュータ及び情報端末

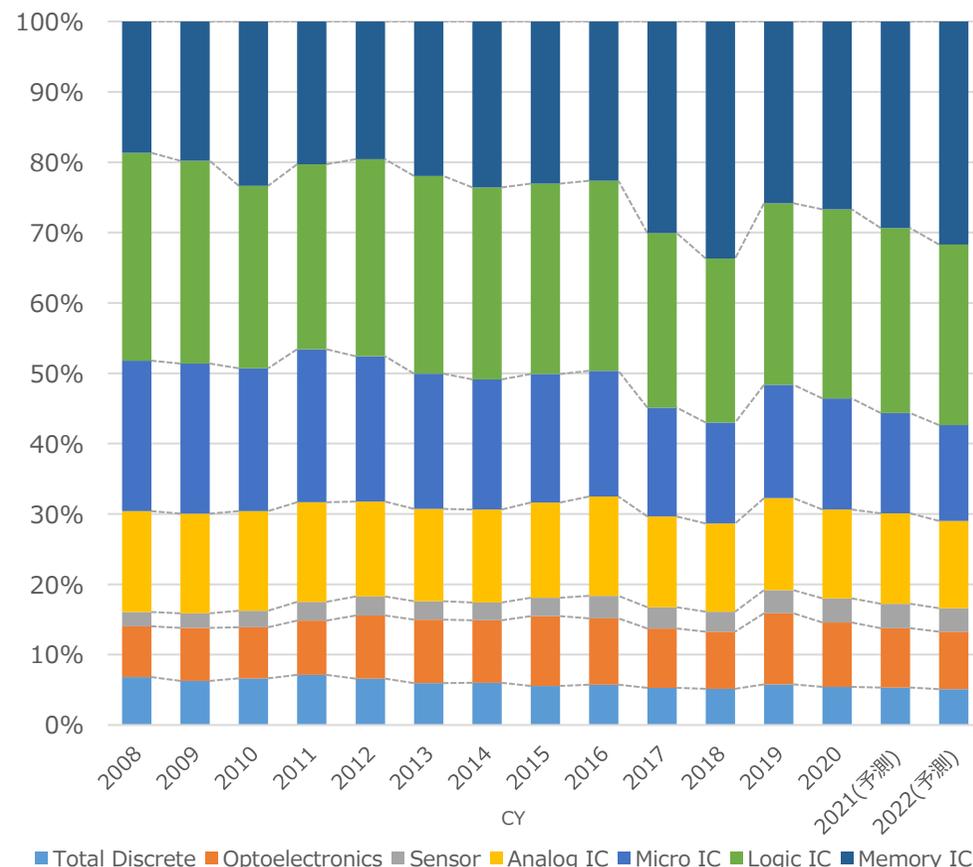
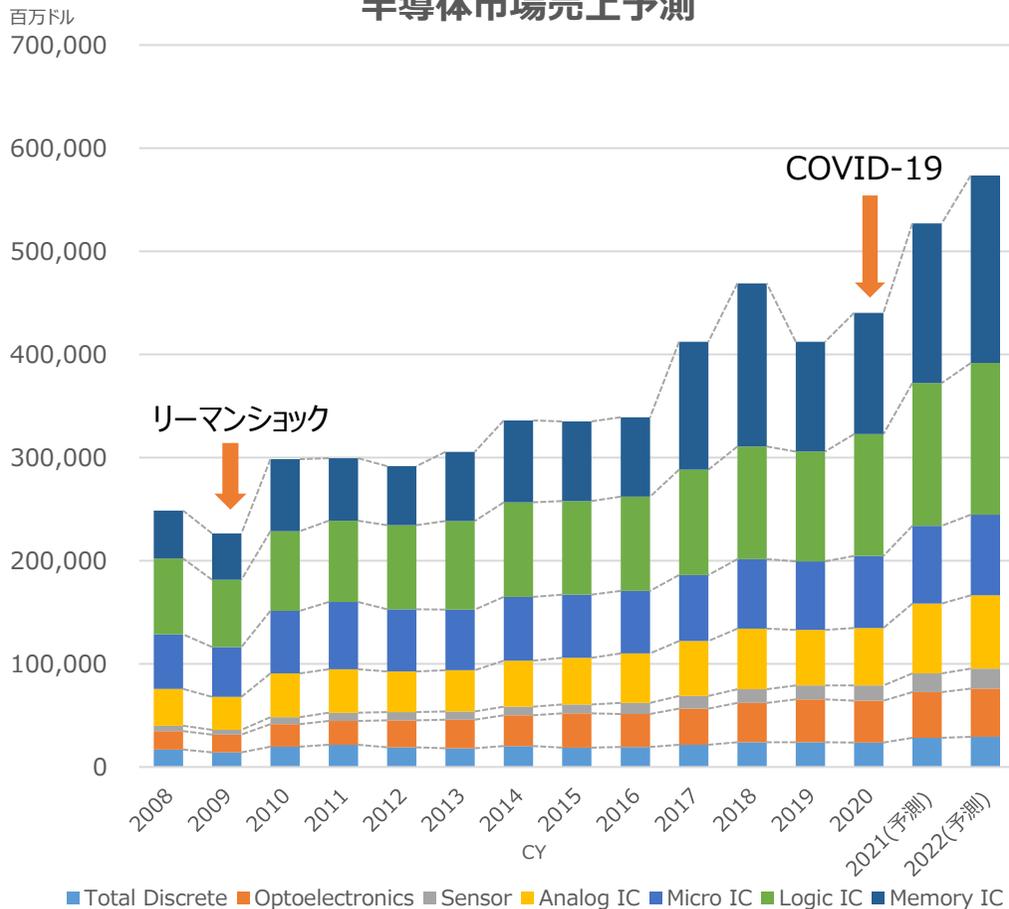
(資料) JEITAの資料をもとにNTTアドバンステクノロジー社が作成

1-1-3. 世界半導体市場売上予測

- 2020年、2021年とCOVID-19禍におけるニューノーマルの生活・仕事スタイルが要因となり、半導体需要は高く成長が続くと見られている。
- 半導体種別では、Memory ICとLogic ICで50%以上を占めている。また、Memory ICは2019年以降に売上比率を伸ばしており、メモリの用途の広がりや需要の堅調さがうかがえる。

半導体市場売上予測

半導体市場売上割合予測



(資料) WSTSの情報をもとにNTTアドバンステクノロジー社が作成

1-2. 主要なマイクロエレクトロニクスに係る製造 メーカ及び製造装置メーカ、材料メーカに係 る市場動向、主要用途分析、技術動向、生 産開発拠点及び生産能力構築状況の調査

1-2-1. 調査の目的と背景

主要なマイクロエレクトロニクスに係る製造メーカー、製造装置・材料メーカーに係る市場動向、主要用途、技術動向、生産開発拠点及び生産能力構築状況について調査する。

【現状】

- COVID-19の影響による半導体製造工場の一時的な閉鎖や、米国の中国への先端半導体技術の規制に端を発する半導体サプライチェーンのデカップリング、また、COVID-19に起因する電子機器の需要増などさまざまな要因により半導体不足が顕著になった。

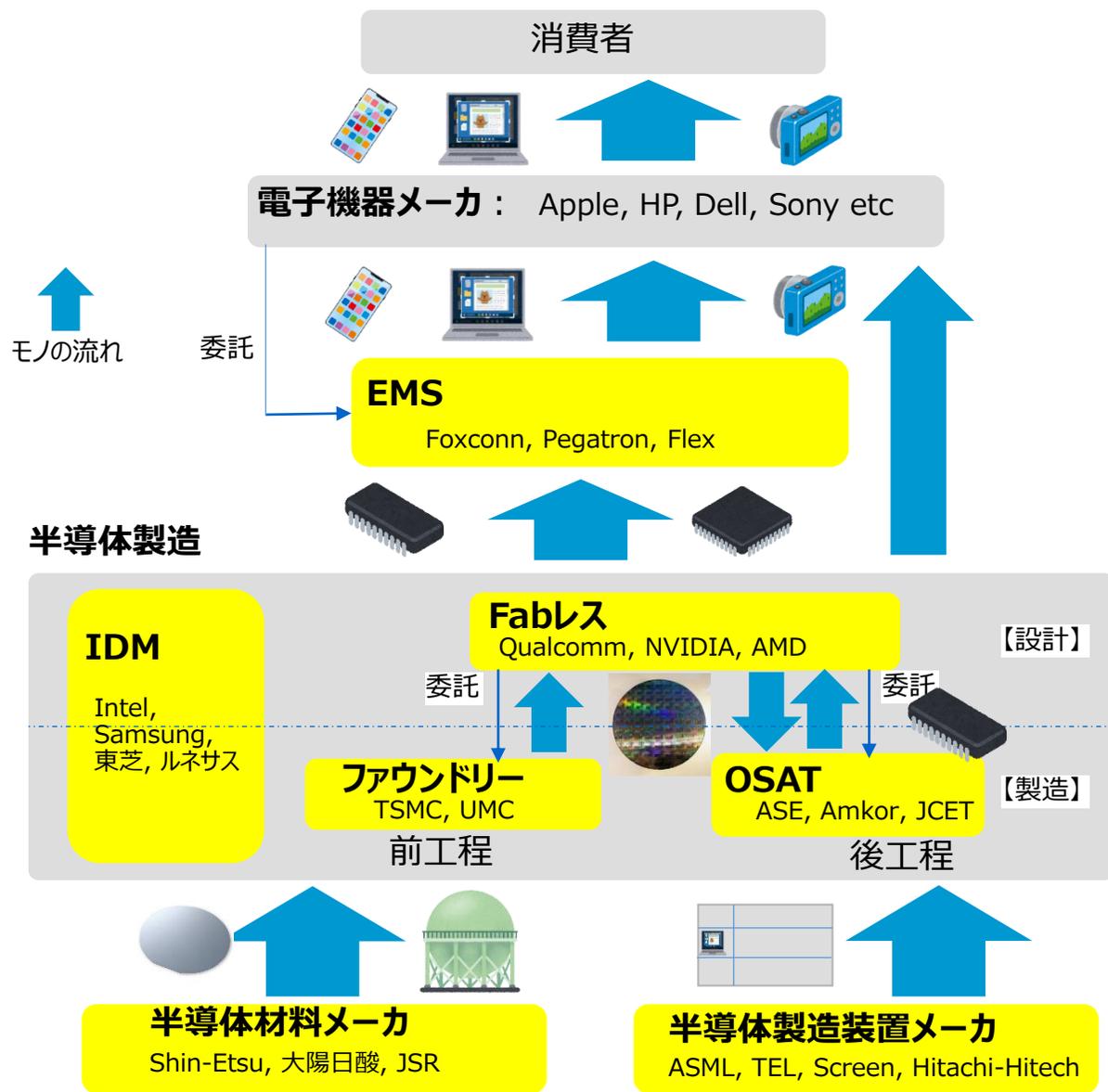
【目的】

- マイクロエレクトロニクスのサプライチェーンを材料から製品までの各市場を調査し、それぞれの市場でのメーカーの動向を把握する。

【調査方法】

- 市場レポート、調査対象個社の年次報告書、およびネットニュース等の公開情報をもとに調査した。
- マイクロエレクトロニクスのサプライチェーンを次スライドに示す。また、次スライドの黄色で示した市場について調査・報告する。

マイクロエレクトロニクスのサプライチェーン



EMS : Electronics Manufacturing Service

電子機器の製造を受託する生産形態・サービスのことを指す。ロット生産を可能とする点が特徴であり、製造だけでなく設計、部品調達や流通なども一貫して行う場合もある。発注元は委託したプロセスについては関与しないことが多い。

IDM : Integrated Device Manufacturer

半導体に関して、設計から販売まで一貫して全行程を行うことができる企業のことを指す。垂直統合型デバイスメーカーともいう。

Fabless : fabless

半導体製造工場を持たず製造プロセスをファウンドリーに委託し、設計や販売を行う企業のことを指す。

ファウンドリー: Foundry

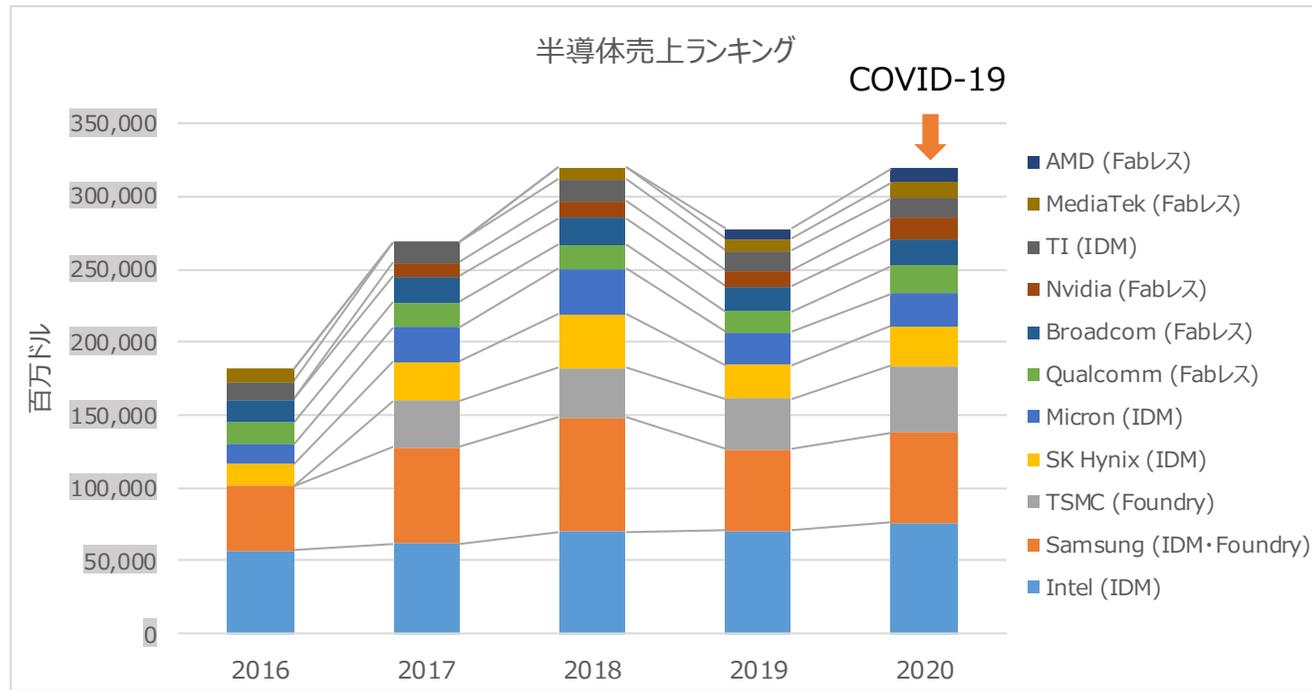
半導体製造のみを受託し、特に半導体製造の前工程の製造を担う企業・工場を指す。ファブレスが設計を行ったものをファウンドリーで製造する形となる。

OSAT : Outsourced Semiconductor Assembly & Test

半導体製造のみを受託し、特に半導体製造の後工程を担う企業のことを指す企業・工場のことを指す。

半導体売上ランキング

- 半導体売上ランキング上位10社の推移(2016年-2020年)を示す。
- トップ2は2016年より変わらず、IntelとSamsungである。
- 近年はTSMCがSamsungの売上高に迫るものとなっている。
- IDMの売上がFabレスよりも大きい傾向にある。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるため0を追加した。

(資料) ICInsightsの情報をもとにNTTアドバンステクノロジーが作成

1-2-2. Foundry市場

1-2-2-1.Foundry上位企業の売上規模ランキング（Top5）

- Foundry上位5社の半導体売上(2016年-2020年)を示す。
- 2016年と比較しても上位5社全て売上が伸びているが、TSMCは圧倒的に伸びていることがわかる。これはTSMCがいち早く最先端プロセスにアクセスし、高い単価で売るため売上が高くなるのだと考えられる。
- SMICの売上が伸びているのは、中国国内の半導体需要の多くを引き受けているためと言われている。SMICの売上はレガシープロセスが中心であるため2020年12月の米国エンティティリスト追加後も影響は少ないと考えられる。
- 2019年は2018年まで続いたスーパーサイクルの反動で一時的に売上が減少している企業が多く、2020年になるとCOVID-19の巣ごもり需要等で売上が2018年よりも増えている企業が多い。

Foundry上位5社の売上高推移（単位：百万ドル）

	2016	2017	2018	2019	2020	2020年/2016年増減率
TSMC(台湾)	29,437	32,040	34,379	34,636	45,967	56%
Samsung(韓国)	10,982	11,527	12,467	12,267	14,471	32%
UMC(台湾)	4,587	4,898	5,073	4,820	5,959	30%
GF(米国)	4,999	5,407	6,202	5,702	5,959	19%
SMIC(中国)	2,914	3,099	3,361	3,115	4,256	46%

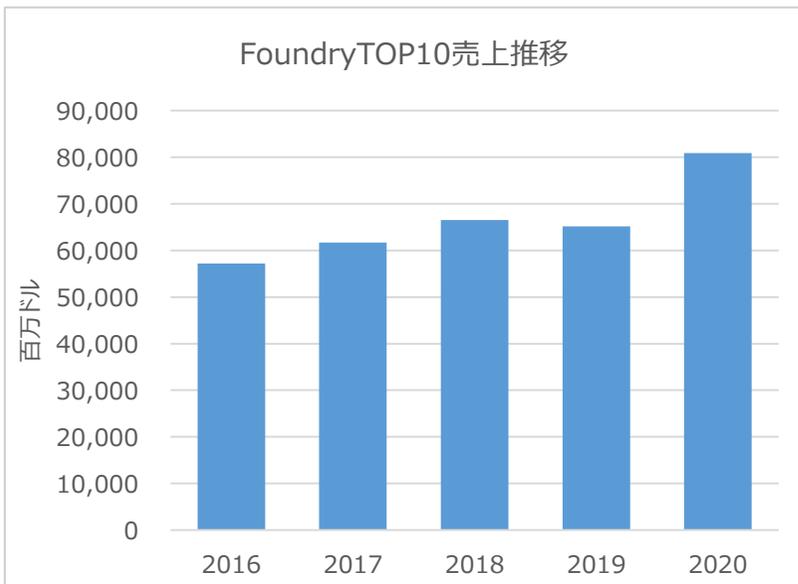
・2018年以降のSamsungの収益にはLSI事業とFoundry事業の収益が含まれている
・Samsungの2016年2017年の数値は2018年と同じ取り扱いにするために、Samsungの公表資料のFoundryの収益を含んでいると思われるLSI事業より抜粋
・GlobalFoundries(GF)の一部年にはIBMから買収したチップ製造ユニットによる収益が含まれている
・2020年はTrendForceで公開されていた全体売上と企業別シェアから売上高を算出した

（資料）TrendForceの情報をもとにNTTアドバンステクノロジーが作成

1-2-2-2. Foundry上位企業の売上規模ランキング (Top10)

- Foundry上位10社の売上(2016-2020年)を示す。
- 5年連続TSMCがトップであり、TOP10トータルの数値と比べても圧倒的な売上を誇る事がわかる。
- TOP10トータルの数値も昨年と比べて増えている。

FoundryTOP10売上推移



Foundry上位10社の売上高推移 (単位：百万ドル)

	2016	2017	2018	2019	2020	20/16増減率
TSMC(台湾)	29,437	32,040	34,379	34,636	45,967	56%
Samsung(韓国)	10,982	11,527	12,467	12,267	14,471	32%
UMC(台湾)	4,587	4,898	5,073	4,820	5,959	30%
GF(米国)	4,999	5,407	6,202	5,702	5,959	19%
SMIC(中国)	2,914	3,099	3,361	3,115	4,256	46%
Towerjazz(米国)	1,249	1,388	1,305	1,234	851	-32%
PSMC(台湾)	870	1,035	1,270	825	851	-2%
VIS(台湾)	801	817	957	916	851	6%
Hua Hong(中国)	721	807	930	933	851	18%
DB HiTek(韓国)	666	676	599	691	851	28%
Top10total	57,226	61,694	66,543	65,139	80,868	41%

・2018年以降のSamsungの収益にはLSI事業とFoundry事業の収益が含まれている
 ・Samsungの2016年2017年の数値は2018年と同じ取り扱いにするために、Samsungの公表資料のFoundryの収益を、Fabと思われるLSI事業より抜粋
 ・GlobalFoundries(GF)の一部年にはチップ製造ユニットによる収益が含まれている
 ・PSMCはFoundry事業のみの収益である。
 ・HuaHongの収益は公表された収益化のみを含んでいる。
 ・2020年はTrendForceの全体売上と企業別シェアから算出した。

(資料) TrendForceの情報をもとにNTTアドバンステクノロジーが作成

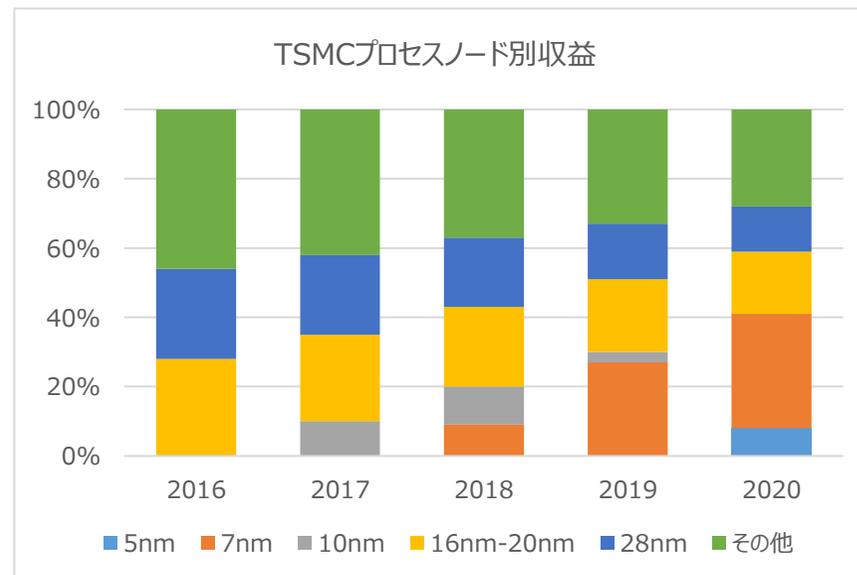
1-2-2-3.TSMC (台湾)

1-2-2-3-1. TSMC プロファイル

- TSMCは台湾に本社を持つFoundry企業売上トップの企業である。
- 2020年度の売上高は2019年度の売上高より30%以上伸びている。市場の中では特に**高性能コンピュータ市場で大きく売上を伸ばして**おり、2019年度の売上高より39%伸びていた。
- 近年7nmプロセスの需要が高く、**2020年度は収益の3割以上が7nmプロセス**である。
- 2020年に5nmプロセスの量産を開始し、**現在3nmプロセスの量産開発に取り組んでいる**。
- 3カ国・地域に12工場を持ち、現在もアメリカのアリゾナ及び台湾の新竹市に300mmウェーハの工場の建設予定または建設中である。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2020年はTrendForceで公開されていた全体売上と企業別シェアから売上高を算出した。



(資料) TrendForce、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-3-2. TSMC Fabロケーション



3カ国・地域に27工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

※"+ "は生産開始に至っていない施設があることを示す

※生産までは至っていないものも含めると42工場

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-3-3. TSMC Fab詳細情報(1/4)

工場名	Fab 2	Fab 3	Fab 5	Fab 6 Phase 1	Fab 6 Phase 2	Fab 8
所在地	Shanghai,China	Camas,Washington, US	Hsinchu,Northern Taiwan,Taiwan	Tainan,Southern Taiwan,Taiwan	Tainan,Southern Taiwan,Taiwan	Taichung,Central Taiwan,Taiwan
テクノロジー	CMOS, GaN-On-Silicon	CMOS, SiGe, BICMOS	CMOS,	CMOS; Cu; Low-K	CMOS; Cu; Low-K	CMOS
Wafer size(inch)	6	8	8	8	8	8
Geometry (microns)	0.450	0.150	0.150	0.110	0.110	0.110
Fab type	Fab	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)						

工場名	Fab 10 (Songjiang)	Fab 11 (WaferTech)	Fab 12 Phase 1&2	Fab 12 Phase 4/5	Fab 12 Phase 6	Fab 12 Phase 7
所在地	Nanjing,Jiangsu,China	Taichung,Central Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan
テクノロジー	CMOS	CMOS, Embedded flash	CMOS, Cu, Low-K, SoC, strained Si	CMOS, Cu, Low-K, SoC, strained Si	CMOS, EUV, Cu, Low-K, SoC, strained Si	EUV, CMOS
Wafer size(inch)	8	8	12	12	12	12
Geometry (microns)	0.130	0.160	0.040	0.005	0.003	0.003
Fab type	Fab	Fab	Fab, R&D	Fab, Pilot, R&D	Fab, Pilot, R&D	Fab, Pilot, R&D
Status(製造中を除く)			In production-merged	In production-merged		

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-3-3. TSMC Fab詳細情報(2/4)

工場名	Fab 12 Phase 8 (was R1)	Fab 12 Phase 9 (was R2)	Fab 14 Phase 1&2	Fab 14 Phase 3	Fab 14 Phase 4	Fab 14 Phase 5
所在地	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan
テクノロジー			CMOS, Cu, Low-K, strained Si	CMOS, Cu, Low K	CMOS	CMOS, Cu, Low-K, SoC, strained Si
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.003	0.002	0.040	0.040	0.040	0.016
Fab type	Fab, Pilot, R&D	Fab, Pilot, R&D	Fab	Fab	Fab	Fab
Status(製造中を除く)	Constructing	Planned	In production-merged			

工場名	Fab 14 Phase 6	Fab 14 Phase 7	Fab 14 Phase 8	Fab 15 Phase 1	Fab 15 Phase 2	Fab 15 Phase 3
所在地	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan
テクノロジー	CMOS, Cu, Low-K, SoC, strained Si, FinFET	CMOS, Cu, Low-K, SoC, strained Si, FinFET	CIS	CMOS, Cu, Low-K, strained Si	CMOS, Cu, Low-K, strained Si	CMOS, Cu, Low-K, strained Si
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.016	0.016	0.040	0.028	0.028	0.028
Fab type	Fab, R&D	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)			Planned			

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-3-3. TSMC Fab詳細情報(3/4)

工場名	Fab 15 Phase 4	Fab 15 Phase 5	Fab 15 Phase 6	Fab 15 Phase 7	Fab 16 P1 (Nanjing)	Fab 16 P2 (Nanjing)
所在地	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan	Hsinchu,NorthernTaiwan,Taiwan
テクノロジー	CMOS, Cu, Low-K, SoC, strained Si	CMOS, FinFET, Cu, Low-K, SoC	CMOS, FinFET, Cu, Low-K, SoC	EUV		
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.022	0.007	0.007	0.007	0.016	0.028
Fab type	Fab	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)						Constructing

工場名	Fab 18 Phase 1	Fab 18 Phase 2	Fab 18 Phase 3	Fab 18 Phase 4	Fab 18 Phase 5	Fab 18 Phase 6
所在地	Tainan,SouthernTaiwan,Taiwan	Tainan,SouthernTaiwan,Taiwan	Tainan,SouthernTaiwan,Taiwan	Tainan,SouthernTaiwan,Taiwan	Tainan,SouthernTaiwan,Taiwan	Tainan,SouthernTaiwan,Taiwan
テクノロジー	CMOS, Cu, Low-K, SoC, strained Si					
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.005	0.005	0.005	0.003	0.003	0.003
Fab type	Fab	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)				Equipping	Constructing	Constructing

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-3-3. TSMC Fab詳細情報(4/4)

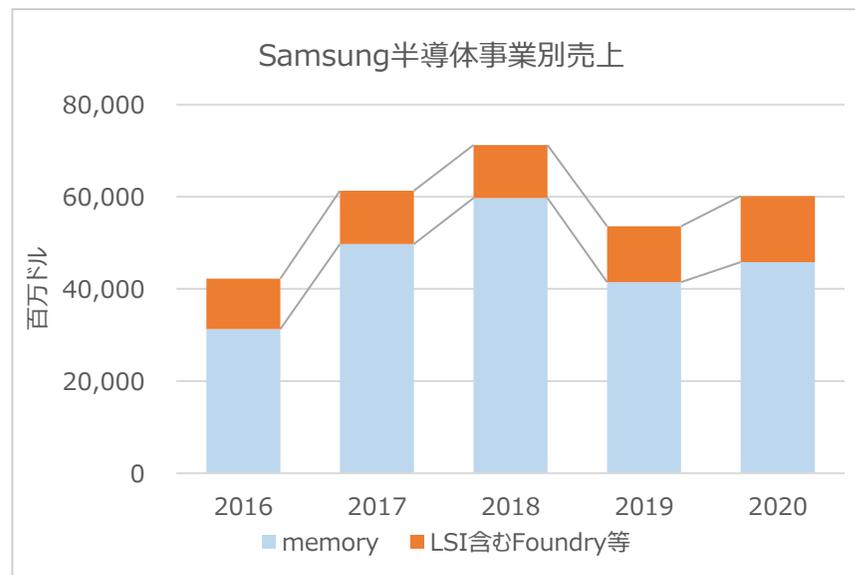
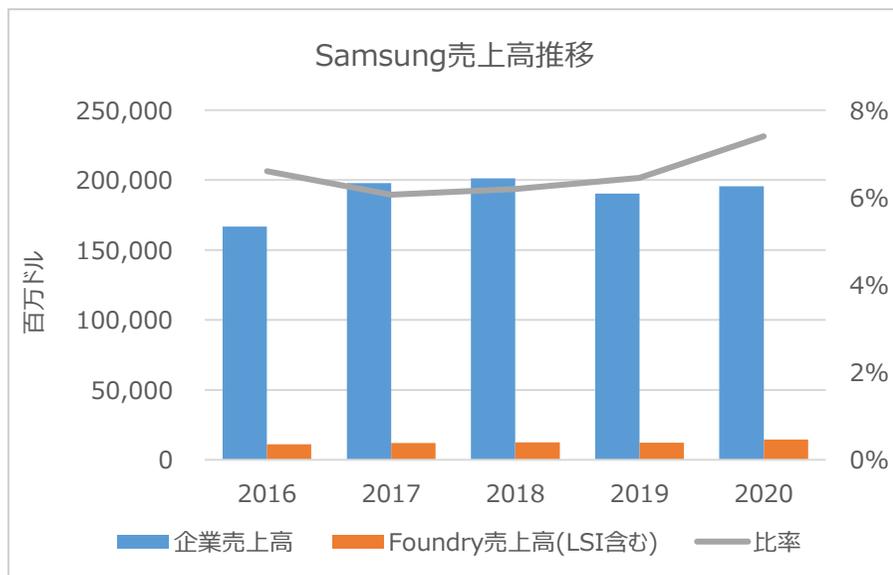
工場名	Fab 18 Phase 7	Fab 18 Phase 8	Fab 21 (Arizona phase 1)	Fab 21 (Arizona phase 2)	N2 fab (P1)	N2 fab (P2)
所在地	Tainan,Southern Taiwan,Taiwan	Tainan,Southern Taiwan,Taiwan	Phoenix,Arizona,Americas	Phoenix,Arizona,Americas	Hsinchu,Northern Taiwan,Taiwan	Hsinchu,Northern Taiwan,Taiwan
テクノロジー						
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.002	0.002	0.005	0.005	0.002	0.002
Fab type	Fab	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)	Planned	Planned	Constructing	Planned	Planned	Planned

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-2-4. Samsung (韓国)

1-2-2-4-1. Samsung プロファイル

- Samsungは韓国に本社を持つFoundry市場売上世界2位の企業である。
- Samsungの半導体事業としてはメモリ事業とシステムLSI事業、Foundry事業がある。
- 2020年度はCOVID-19の影響でIT機器の需要が増加したことにより、Foundry事業の売上が増加している。
- 2018年にEUVリソグラフィ技術による7nmプロセスの生産を開始し、現在は5nmプロセスも量産している。
- **2022年後半にGAA(Gate-All-Around)プロセスを採用した3nmプロセスノードの商業生産を開始予定。**
- 2カ国・地域に17工場を持つ。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



Foundry売上高(LSI含む)のグラフについて

- ・2018年以降のSamsungの収益にはLSI事業とFoundry事業の収益が含まれている
- ・Samsungの2016年2017年の数値は2018年と同じ取り扱いにするために、Samsungの公表資料のFoundryの収益を含んでいると思われるLSI事業より抜粋
- ・2020年はTrendForceで公開されていた全体売上と企業別シェアから算出

・十億単位であったため、100万ドル単位に合わせた

(資料) TrendForce、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成
 企業売上高及び半導体事業別売上高の数値は2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

1-2-2-4-2. Samsung Fabロケーション



2カ国・地域に15工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

※“+”は生産開始に至っていない施設があることを示す。

※製造開始までは至っていないものも含めると17工場

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-4-3. Samsung Fab詳細情報(1/2)

WFFにおいてFoundry/IDMカテゴリに含まれていたファブを抜粋。

工場名	Copper fab (Austin 1, BEOL)	Line 6	Line 7	Line 8 (200mm)	Line 13 (CIS Line)	P3 Phase 2 (Logic 8th floor)
所在地	Austin,Texas,US	Giheung,Gyeonggi,Korea	Giheung,Gyeonggi,Korea	Giheung,Gyeonggi,Korea	Hwaseong,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea
テクノロジー	CMOS	CMOS, Cu, CIS	CMOS, Cu, SOI	CMOS		
Wafer size(inch)	12	8	8	8	12	12
Geometry (microns)	0.090	0.065	0.065	0.065	0.014	0.003
FabType	Fab, BEOL	Fab	Fab	Fab	Fab/Split	Fab
Status(製造中を除く)					Planned	Cancelled

工場名	S? New Austin Fab	S1 (CIS Line)	S1 (former S1-A Line includes S1-C start 1Q16)	S1-B Line BEOL (former Line 9 & Line 8)	S2 (Main Fab)	S3-3rd floor (part of Line 17) EUV
所在地	Austin,Texas,US	Giheung,Gyeonggi,Korea	Giheung,Gyeonggi,Korea	Giheung,Gyeonggi,Korea	Austin,Texas,US	Hwaseong,Gyeonggi,Korea
テクノロジー	CMOS, Cu, SOI, strained Si, low-k, high-k/metal gate, FDSOI	CIS	CMOS, Cu, SOI, strained Si, low-k, high-k/metal gate, FDSOI	CMOS, Cu	CMOS, Cu, SOI, strained Si, low-k, high-k/metal gate, FDSOI	FinFET, EUV
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.003	0.028	0.008	0.028	0.011	0.005
FabType	Fab	Fab/Split	Fab	Fab, BEOL	Fab	Fab
Status(製造中を除く)	Planned				In production-merged	

※グレーの網掛けは中止されたFab

※工場名に記載した?は番号未定のため便宜的に記載している

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-4-3. Samsung Fab詳細情報(2/2)

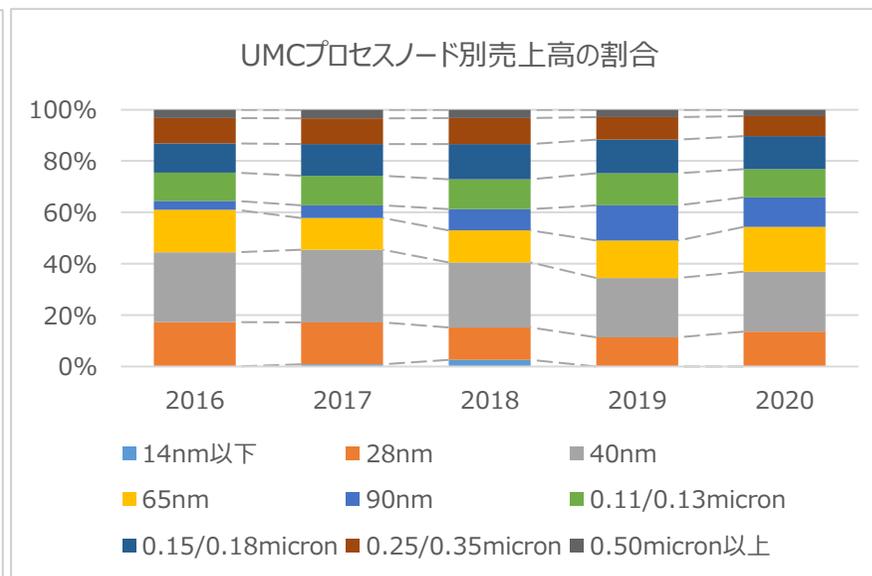
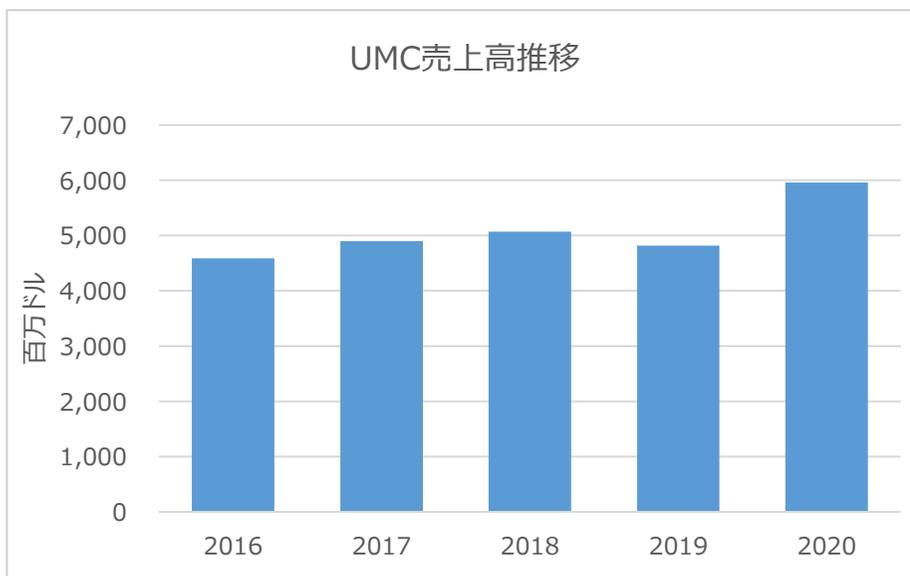
工場名	S4 CIS (part of Line 11)	S5 (part of P2, Phase 1, 3rd floor)	V1- EUV 1st floor (was EUV Fab)	V1-Logic 2nd floor (was EUV Fab)	V2 (EUV) 1st floor (EUV Fab)	V2 (Logic) 2nd floor
所在地	Hwaseong,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea	Hwaseong,Gyeonggi,Korea	Hwaseong,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea
テクノロジー	CMOS	CMOS	EUV	Logic	Logic and DRAM	
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.028	0.005	0.005	0.005	0.005	0.003
FabType	Fab	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)						Planned

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-5. UMC (台湾)

1-2-2-5-1. UMC プロファイル

- UMCは台湾に本社を持つ半導体製造企業である。
- 1995年にIDMから専業Foundryに転換した。2020年の売上においてはFoundry企業の中で第三位となっている。
- 2020年は在宅勤務や遠隔授業により半導体の需要が増加したため売上が増加した。
- ノード別売上高の割合としては40nmプロセスの割合が最も多いが、近年65nmプロセスの割合も増えている。2019年から2020年にかけては14nm以下の収益はない。
- 2019年に日本のUSJC社を買収し、現在日本の工場も含め4カ国・地域に12工場を持つ。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2020年はTrendForceの全体売上と企業別シェアから算出した。

(資料) TrendForce、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-5-2. UMC Fabロケーション



(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-5-3. UMC Fab詳細情報(1/2)

工場名	Fab 6A (Wavetek)	Fab 8A	Fab 8C	Fab 8D	Fab 8E
所在地	Hsinchu,Northern Taiwan,Taiwan				
テクノロジー	CMOS, Al, FSG	CMOS, Al, FSG	CMOS, Al, FSG	CMOS, Cu, FSG, low K	CMOS, Al, FSG
Wafer size(inch)	6	8	8	8	8
Geometry (microns)	0.150	0.180	0.110	0.090	0.130
FabType	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)					

工場名	Fab 8F	Fab 8N (former HeJian)	Fab 8S	Fab 12A P1/P2	Fab 12A P3/P4
所在地	Hsinchu,Northern Taiwan,Taiwan	Suzhou,Jiangsu,China	Hsinchu,Northern Taiwan,Taiwan	Tainan,Southern Taiwan,Taiwan	Tainan,Southern Taiwan,Taiwan
テクノロジー	CMOS, Al, FSG	CMOS, Cu	CMOS, Al, FSG	CMOS, Cu, Al, FSG, low K, SoC	CMOS, Cu
Wafer size(inch)	8	8	8	12	12
Geometry (microns)	0.110	0.110	0.110	0.040	0.022
FabType	Fab	Fab	Fab	Fab, R&D	Fab
Status(製造中を除く)		In production-merged			

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-2-5-3. UMC Fab詳細情報(2/2)

工場名	Fab 12A P5/P6	Fab 12I	Fab 12M B1 (was Fujitsu Mie Fab)	Fab 12M B2 (was Fujitsu Mie Fab)	Fab 12X
所在地	Tainan,Southern Taiwan,Taiwan	Singapore,Singapore,Singapore	Kuwana,Mie,Japan	Kuwana,Mie,Japan	Xiamen,Fujian,China
テクノロジー	CMOS, Cu, Low-K, SoC, strained Si	CMOS, Cu, FSG, Low-K, SoC	CMOS, Cu	CMOS; Cu; low-K, Low power	CMOS, Cu
Wafer size(inch)	12	12	12	12	12
Geometry (microns)	0.028	0.040	0.090	0.040	0.028
FabType	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)					

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-2-6.GF (米国)

1-2-2-6-1. GFプロフィール

- GlobalFoundries(GF)はアメリカに本社を置く半導体製造企業であり、アラブ首長国連邦アブダビ首長国の開発公社が出資している企業である。
- 売上の大部分を占めるのはスマートフォン市場向け製品である。近年は自動車関連需要が増加しており、BMW社向けのチップ開発をしているINOVAsemiconductors社と半導体供給契約を締結しBMWと半導体に関する提携を結ぶことや、Ford社と米国内での半導体製造と技術開発を促進するための提携を結んだことを2021年に発表している。
- 2021年2月、GFは米国防総省(DoD)との間に戦略的パートナーシップを締結し、ニューヨーク州のFab 8で国家安全保障上の機密性の高い半導体チップを製造することを発表。45nm SOIプラットフォーム上で製造する。
- 2021年、ナスダックにIPO申請をしている。
- 3か国・地域に5工場を持ち、現在も米国に12インチのファブ建設を計画中である。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2020年はTrendForceの全体売上と企業別シェアから算出した。

(資料) TrendForceの情報をもとにN T Tアドバンステクノロジーが作成

1-2-2-6-2. GF Fabロケーション



3カ国・地域に8工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

※" +"は生産開始に至っていない施設があることを示す。

※製造開始までは至っていないものも含めると9工場

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-6-3. GF Fab詳細情報

工場名	Fab 1 Mod 1+2 (was mod 2 was Fab 30/38)	Fab 1, Module 3 (Fab 1, Annex)	Fab7G (Fab 7 + Fab 6)	Fab 8 phase 1 (was called Fab 2)	Fab 8 phase 3 (TDC Technology Development Center)
所在地	Dresden,Saxony,Germany	Dresden,Saxony,Germany	Singapore,Singapore,Singapore	Malta,New York,US	Malta,New York,US
テクノロジー	CMOS, Cu, SOI, low-K; Strained Si, FDSOI (FDX)	CMOS, Cu, SOI, low-K; Strained Si; NiSix	CMOS; RF CMOS; RF-SOI; Cu; low-K; high-K, strained Si, MEMS	CMOS, Cu, SOI, low-K; Strained Si; NiSix; Ultra-Low K	FinFET
Wafer size(inch)	12	12	12	12	12
Geometry (microns)	0.022	0.022	0.040	0.012	0.012
FabType	Fab, R&D	Fab	Fab	Fab	Fab, Pilot
Status(製造中を除く)	In production-merged		In production-merged	In production-merged	

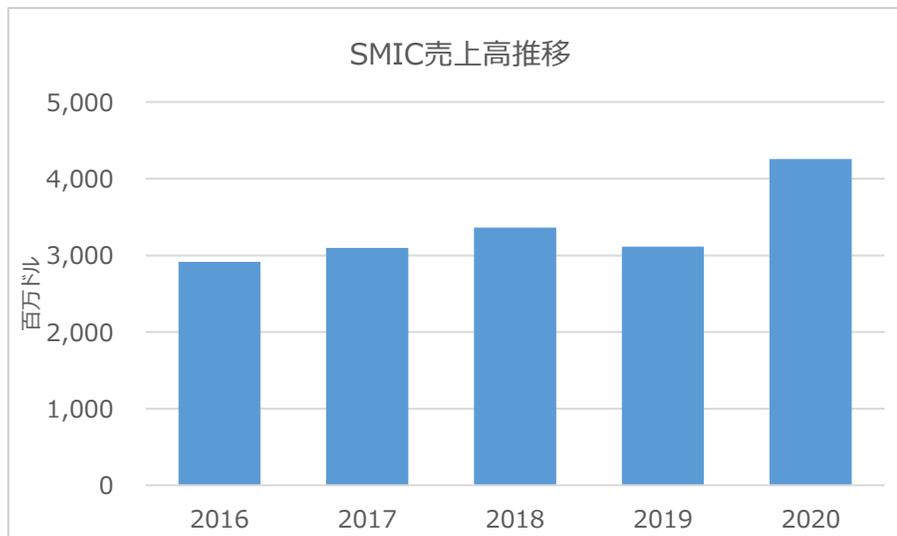
工場名	Fab 8.2	Fab 9 (was IBM B 973 & B 970)	Fab 10 (was Glofo Fab 10, was IBM Bldg. 323 plus Annex)	Giga Fab (Fab 2,3,5, was Fab 2)
所在地	Malta,New York,US	Essex Junction,Vermont,US	East Fishkill,New York,US	Singapore,Singapore,Singapore
テクノロジー		CMOS; BiCMOS; Cu; SiGe; GaN; SOI;	CMOS, BiCMOS, Cu, FD-SOI, SiGe, ultra low-K, high-K, strained Si, Airgaps, FDSOI (FDX)	CMOS; BiCMOS; HVCMOS, AI
Wafer size(inch)	12	8	12	8
Geometry (microns)	0.010	0.090	0.022	0.180
FabType	Fab	Fab	Fab, Pilot, R&D	Fab
Status(製造中を除く)	Planned	In production-merged		In production-merged

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

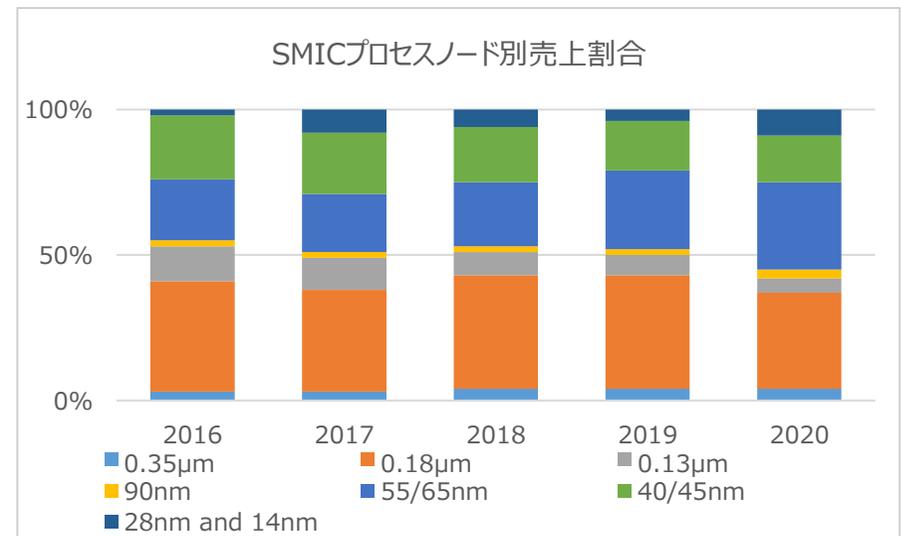
1-2-2-7. SMIC (中国)

1-2-2-7-1. SMIC プロファイル

- SMICは中国上海に本社を置く半導体製造企業である。
- 2019年末にFinFETプロセスのリスク生産を開始し、2020年度には月に15,000ウェーハの生産能力を有している。
- 350nmから14nmまでを扱い、**主力となるのは180nmと55/65nm**だが、近年は14nmの売上の割合も増えてきている。
- **地理的な売上高としては中国や香港が全体の6割を占める。**
- 2020年にSMIC社は米国の輸出規制の対象となっているが、**中国国内の旺盛な半導体需要を引き受けて売上を伸ばしている。**
- 1カ国・地域に13工場を持つ。北京に12インチのFabを建設中であり、また深センにも12インチの工場建設を計画している。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2020年はTrendForceの全体売上と企業別シェアから算出した。



(資料) TrendForce、企業HP、年度報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-7-2. SMIC Fabロケーション



1カ国・地域に14工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

※"+ "は生産開始に至っていない施設があることを示す。

※製造開始までは至っていないものも含めると20工場

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-7-3. SMIC Fab詳細情報(1/2)

工場名	B1 MegaFab (Fab 4)	B1 MegaFab (Fab 5)	B2	B3	Fab 7
所在地	Beijing, Beijing, China	Beijing, Beijing, China	Beijing, Beijing, China	Beijing, Beijing, China	Tianjin, Tianjin, China
テクノロジー	CMOS, BiCMOS, BiPolar; Cu, low-K	CMOS, BiCMOS, BiPolar, Cu, low-K	CMOS, Cu, low-K	CMOS, Cu, low-K	BICMOS, CMOS, AI
Wafer size(inch)	12	12	12	12	8
Geometry (microns)	0.040	0.040	0.028	0.028	0.150
FabType	Fab, R&D	Fab	Fab	Fab	Fab
Status(製造中を除く)					

工場名	Fab 8	Fab 15 (Shenzhen 8-inch fab)	Fab 16 (Shenzhen 12-inch fab)	MegaFab (Fab 3 B-C)	MegaFab Fab 6
所在地	Shanghai, Shanghai, China	Shenzhen, Guangdong, China	Shenzhen, Guangdong, China	Shanghai, Shanghai, China	Beijing, Beijing, China
テクノロジー	CMOS, BiCMOS, Cu, low-K	CMOS	CMOS	CMOS, Cu	Metalization; CMOS, Cu, AI
Wafer size(inch)	12	8	12	8	12
Geometry (microns)	0.014	0.130	0.040	0.130	0.090
FabType	Fab, R&D	Fab	Fab	Met	Met
Status(製造中を除く)	Closed as semi fab but continues		Planned		production/support

※グレーの網掛けは中止されたFab

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-2-7-3. SMIC Fab詳細情報(2/2)

工場名	New Beijing JV fab	Ningbo N1	Ningbo N2	S1 MegaFab (Fab 1)	S1 MegaFab (Fab 2)
所在地	Beijing,China	Ningbo,Zhejiang,China	Ningbo,Zhejiang,China	Shanghai,Shanghai,China	Shanghai,Shanghai,China
テクノロジー		High-voltage BCD, pHEMT	SOI, HV, MEMS, Power	CMOS, FSG as insulator	CMOS, FSG as insulator
Wafer size(inch)	12	8	8	8	8
Geometry (microns)	0.028	0.130	0.250	0.110	0.110
FabType	Fab	Fab	Fab	Fab	Fab
Status(製造中を除く)	Constructing		Constructing		

工場名	Shaoxing Compound fab	Shaoxing JV fab	SN1	SN2	Tianjin fab (T2/T3)
所在地	Shaoxing,Zhejiang,China	Shaoxing,Zhejiang,China	Shanghai,Shanghai,China	Shanghai,Shanghai,China	Tianjin,Tianjin,China
テクノロジー					
Wafer size(inch)	6	8	12	12	8
Geometry (microns)	1.000	0.500	0.014	0.014	0.130
FabType	Fab	Fab	Fab, Pilot, R&D	Fab	Fab
Status(製造中を除く)	Announced			Planned	

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-3. IDM市場

1-2-3-1. IDM上位企業の売上規模ランキング (Top5)

- IDM上位5社の半導体売上(2016年-2020年)を示す。
- SK HynixはTOP3に入っているが、増減率が最も高い。
- 半導体は一定周期で好況と不況を繰り返すが(シリコン・サイクル)、2018年まではメモリ需要が特に増え続けるいわゆるスーパーサイクルに入っていたことから、メモリを扱うSamsung・SK Hynix・Micronの売上高は高いものとなっている。また、NANDよりもDRAMが特に増えており、SK Hynixの伸びが大きくなっている。これは、ビッグデータの普及によりサーバ需要が急拡大し、DRAM需要も伸びたためと言われている。
- TIはTOP5の中で増減率が最も低く、TOP5の中で唯一昨年度より売上が低い。米中貿易摩擦により組込プロセッシング需要が落ちていることが大きな要因の一つであると考えられる。

IDM上位5社の売上高推移 (単位：百万ドル)

	2016	2017	2018	2019	2020	2020年/2016年増減率
Intel(米国)	57,000	61,720	69,880	70,797	76,328	34%
Samsung(韓国)	44,300	65,882	78,541	55,709	61,853	40%
SK Hynix(韓国)	14,900	26,722	36,767	23,185	27,075	82%
Micron(米国)	13,500	23,920	30,930	22,405	22,542	67%
TI(米国)	12,500	13,910	14,854	13,651	13,574	9%

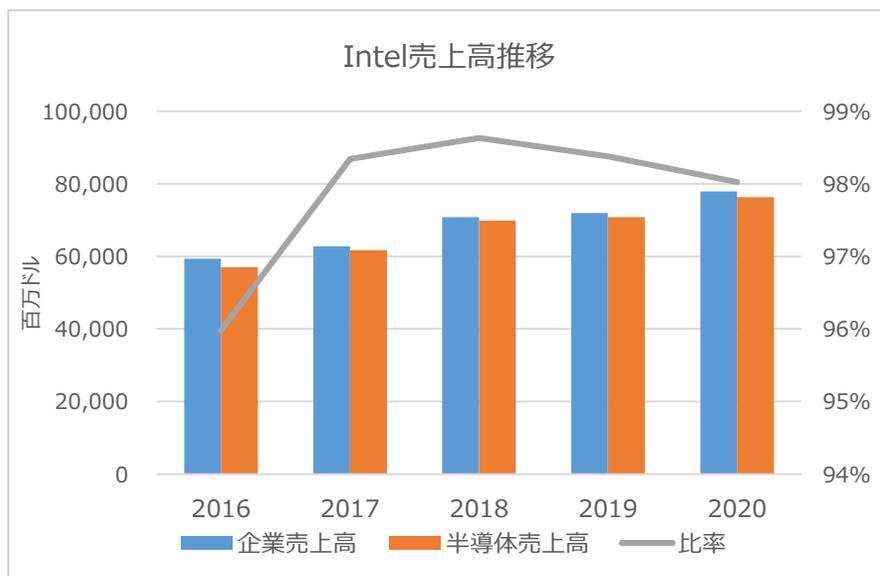
・2016年は十億単位で公表されていたため百万ドル単位に合わせるため0を追加した

(資料) ICInsightsの情報をもとにNTTアドバンステクノロジーが作成

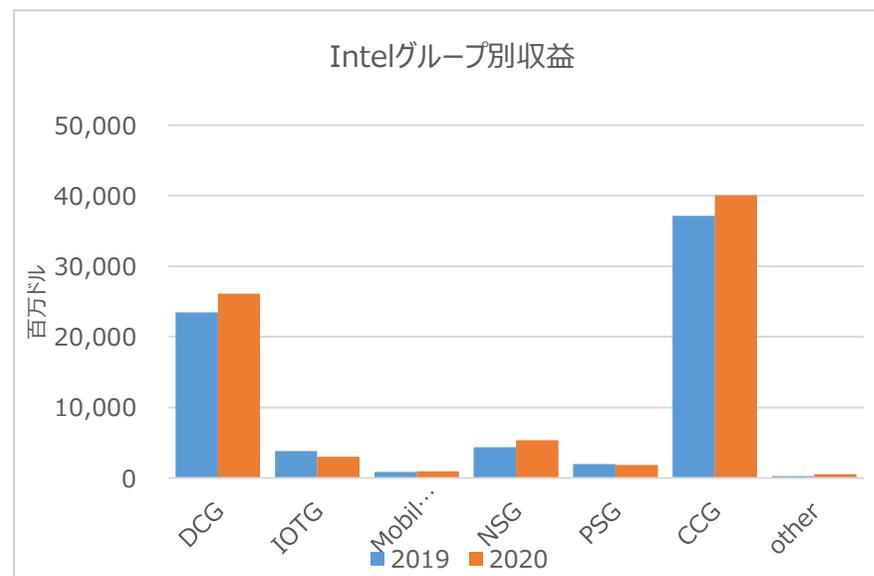
1-2-3-2.Intel (米国)

1-2-3-2-1. Intel プロファイル

- Intelは米国に本社を持つIDM企業の中で売上高トップの企業である。
- 主にフラッシュメモリやマイクロプロセッサの設計開発等を行っている。
- 2020年度半導体に関連する事業のうちData Center Group (DCG), Client Computing Group (CCG), Non-Volatile Memory Solutions Group (NSG)が過去最高の収益を達成している。COVID-19によりノートPCの需要が拡大したことなどが要因と考えられる。
- 2021年には「**IDM2.0**」という事業戦略を発表しており、その中での取り組みの一つである受託製造開始のためのFoundry部門が2022年度より設置される。
- Intelは10カ国・地域で15の製造工場を有し、後工程に関しては6カ国・地域に工場を有する。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるための0を追加した。



(資料) ICInsights、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-2-3. Intel Fabロケーション



(資料) World Fab Forecast(SEMI)、企業HP、年次報告書等、ニュースサイトをもとにNTTアドバンステクノロジーが作成

1-2-3-2-4. Intel Fab詳細情報(1/3)

	【Memory】			【Micro】			
工場名	Fab 11X 3D Xpoint Pilot	Fab 68 (Phase 1)	Fab 88 (was called Fab 68 Phase 2)	D1C (+ D1C expansion Fab 20)	D1D	D1X Module 1	D1X Module 2
所在地	Rio Rancho (Albuquerque), New Mexico, US	Dalian, Liaoning, China	Dalian, Liaoning, China	Hillsboro, Oregon, US	Hillsboro, Oregon, US	Hillsboro, Oregon, US	Hillsboro, Oregon, US
テクノロジー	3D Xpoint, CMOS, Cu, low-k, high-K	CMOS, Cu, low K, strained Si	3D-Xpoint, 3D NAND	CMOS, SiGe, Cu, Low-K, strained Si	CMOS, Cu, Low-K Interconnct, strained Si, High-k, Metal Gate, Immersion Litho	CMOS, FINFET	CMOS, FinFET
Wafer size(inch)	12	12	12	12	12	12	12
Geometry (microns)	0.02	0.02	0.02	0.01	0.01	0.007	0.007
3DLayer	4	144	144				
FabType	R&D, Pilot	Fab, Pilot	Fab	Fab, Pilot, R&D	Fab, Pilot, R&D	Fab, Pilot, R&D	Fab, Pilot, R&D
Status(製造中を除く)				In production-merged			

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-2-4. Intel Fab詳細情報(2/3)

工場名	D1X Module 3	Fab 10 (was IFO)	Fab 11X (+expansion)	Fab 24 (+ Fab 14 combined)	Fab 24 (+Fab14) expansion	Fab 28 (+Fab 28A)	Fab 28A 12-inch (was Fab 18, was Fab 1 Numonyx, Fab 18 Intel)
所在地	Hillsboro,Oregon,US	Leixlip,Kildare,Ireland	Rio Rancho (Albuquerque),New Mexico,US	Leixlip,Kildare,Ireland	Leixlip,Kildare,Ireland	Kiryat Gat,Southern District,Israel	Kiryat Gat,Southern District,Israel
テクノロジー	CMOS, FinFET	CMOS, Cu	CMOS, Cu, Low-K Interconnct, strained Si, High-k, Metal Gate, Immersion Litho	CMOS, Cu, low K, strained Si	CMOS, Cu, low K, strained Si	CMOS, strained Si, Cu, Low-K	CMOS, Cu, low-K, high-K
Wafer size(inch)	12	12	12	12	12	12	12
Geometry (microns)	0.007	0.014	0.032	0.014	0.014	0.01	0.01
3DLayer							
FabType	Fab, Pilot, R&D	Back End was Fab	Fab	Fab, R&D	Fab	Fab	Fab
Status(製造中を除く)	Equipping			In production-merged	Merged	In production-merged	Merged

※オレンジの網掛けは後工程ファブ

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-2-4. Intel Fab詳細情報(3/3)

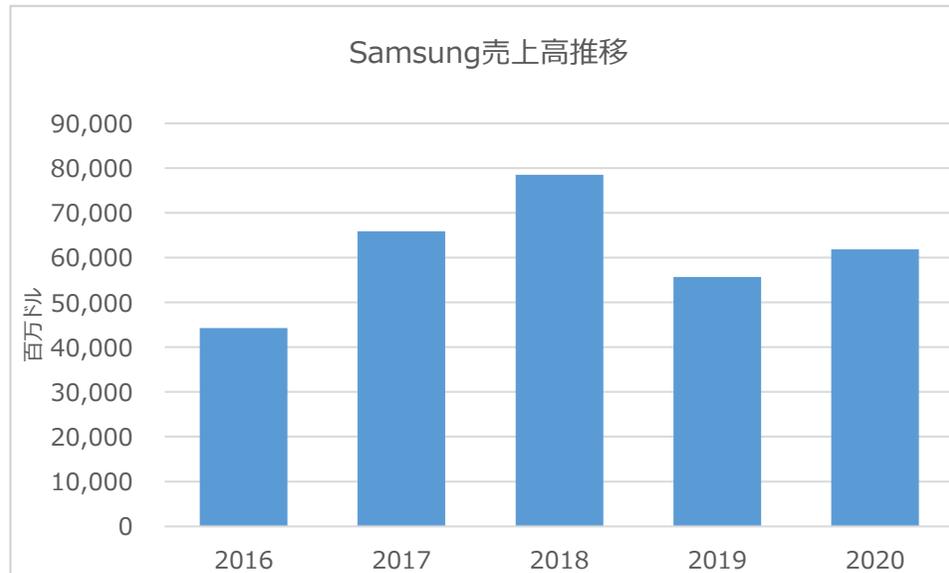
	【Foundry】						【etc】
工場名	Fab 34 (Phase 1 + 2)	Fab 38 Phase 1 + 2 (new Fab Israel)	Fab 42	Megafab (Fab 32+Fab 22+Fab 12)	Fab 52 (estimate)	Fab 62 (estimate)	RP1
所在地	Leixlip, Kildare, Ireland	Kiryat Gat, Southern District, Israel	Chandler, Arizona, US	Chandler, Arizona, US	Chandler, Arizona, US	Chandler, Arizona, US	Hillsboro, Oregon, US
テクノロジー	CMOS	CMOS, strained Si, Cu, Low-K	CMOS, FinFET	CMOS, Cu, Hf-based high K	CMOS, FinFET	CMOS, FinFET	CMOS Cu, low K, SOI, Flash; strained Si
Wafer size (inch)	12	12	12	12	12	12	12
Geometry (microns)	0.007	0.005	0.007	0.014	0.007	0.005	0.007
3D Layer							
Fab Type	Fab	Fab	Fab	Fab	Fab	Fab	R&D
Status (製造中を除く)	Constructing	Announced		In production-merged	Announced	Announced	

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3. Samsung (韓国)

1-2-3-3-1. Samsung プロファイル

- Samsungの半導体分野の売上の割合の大部分がメモリ関係である。その中でDRAMの市場シェアは40%を超えている。
- 2018年まではメモリがスーパーサイクルに入っていたため、売上が増加している。
- 2020年サーバ用DRAM需要が高まっていたこと、また米中貿易摩擦の影響でHuaweiが米国の貿易規制対象にされたことにより、システムLSI関連のモバイル製品の販売が増加したことにより売上が増加した。
- **韓国の平沢に新たにP3ラインの工場を建設予定であり、世界最大規模になるという。**
- 工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるための0を追加した。

(資料) ICInsights、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3-2. Samsung Fabロケーション



2カ国・地域に工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(WFFに情報があつたもののみ記載)

※“+”は製造開始に至っていない工場を示す。

※後工程のファブ関連の情報があつた地域は赤字で示す。

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3-3. Samsung Fab詳細情報(1/4)

【Memory】

工場名	Line 10	Line 12 (2D NAND) + R&D	Line 12 (3D NAND)	Line 13 (DRAM, FEOL)	Line 15-2F & Line 15-1F (8th floor)	Line 15-Cu processing	Line 16 (3D-NAND)
所在地	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea	Hwaseong,Gyeong gi,Korea
テクノロジー	CMOS, Cu	CMOS, Cu	CMOS	CMOS, Cu, low-K	CMOS, Cu, low-K,	Cu	CMOS, Cu, low K, HKMG
Wafer size(inch)	12	12	12	12	12	12	12
Geometry (microns)	0.068	0.014	0.021	0.018	0.017		0.021
3DLayer			96				64
FabType	Fab, BEOL	Fab, R&D	Fab/Split	Fab	Fab	Fab, BEOL	Fab/Split
Status(製造 中を除く)			Cancelled				

※グレーの網掛けは中止されたFab

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-3-3-3. Samsung Fab詳細情報(2/4)

工場名	Line 16 (DRAM)	Line 17-3D NAND MEOL/BEOL (3rd floor) (was called Line 16-2, was called S3-9th floor)	Line 17-DRAM (8th floor+3rd floor)	P1 Phase 1 3D NAND (3rd floor)	P1 Phase 2 3D NAND (8th floor)	P1 Phase 2 DRAM (8th floor)	P2 Phase 1 (DRAM)
所在地	Hwaseong,Gyeonggi,Korea	Hwaseong,Gyeonggi,Korea	Hwaseong,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea	Pyeongtaek,Gyeonggi,Korea
テクノロジー	CMOS, Cu, low K, HKMG	MEOL, BEOL	CMOS	CMOS	CMOS		
Wafer size(inch)	12	12	12	12	12	12	12
Geometry (microns)	0.017	0.01	0.015	0.021	0.021	0.015	0.015
3DLayer		64		128	128		
FabType	Fab/Split	Fab, Pilot,R&D/Split	Fab, Pilot	Fab	Fab	Fab	Fab
Status(製造中を除く)							

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3-3. Samsung Fab詳細情報(3/4)

工場名	P2 Phase 2 (3D NAND)	P3 Phase 1 (3D NAND)	P3 Phase 2 (DRAM)	P4 Phase 1 (3D NAND)	P4 Phase 1 (DRAM)	R Square	SCS Line 1 (former Xian Module 1)
所在地	Pyeongtaek, Gyeonggi, Korea	Hwaseong, Gyeonggi, Korea	Xian, Shaanxi, China				
テクノロジー	3D NAND	3D NAND	DRAM	3D NAND	DRAM	CMOS, Cu, low-K, SOI	CMOS, Cu, low K, HKMG, 3D NAND
Wafer size(inch)	12	12	12	12	12	12	12
Geometry (microns)	0.021	0.021	0.012	0.021	0.011	0.012	0.021
3DLayer	176	238		352			92
FabType	Fab	Fab	Fab	Fab	Fab	R&D	Fab
Status(製造中を除く)		Constructing	Constructing	Not announced	Not announced		

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3-3. Samsung Fab詳細情報(4/4)

	【Opt】			【etc】			
工場名	SCS Line 2 (former Xian Module 2)	Line 4 (LED, was System LSI)	Line 5 (LED line) 8-inch	Line 5 (LED was System LSI)	NRD Line	Xian R&D Center	San Jose
所在地	Xian, Shaanxi, China	Giheung, Gyeonggi, Korea	Giheung, Gyeonggi, Korea	Giheung, Gyeonggi, Korea	Hwaseong, Gyeonggi, Korea	Xian, Shaanxi, China	San Jose, California, US
テクノロジー	3D NAND	III-V	GaN	CMOS, LED	PRAM, FRAM, DRAM, Flash; CMOS		
Wafer size (inch)	12	4	8	4	12	12	12
Geometry (microns)	0.021	2	2	1	0.01	0.028	0.01
3D Layer	92						
Fab Type	Fab	Fab/EPI	Fab/EPI/Split	Fab/EPI	R&D	R&D	R&D
Status (製造中を除く)							

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-3-4. Samsung 後工程Fab詳細情報

工場名	TSV (was N Line)
所在地	Giheung,Gyeonggi,Korea
テクノロジー	CMOS, Cu, low-K, SOI
Wafer size(inch)	8
Geometry (microns)	0.032
3DLayer	
FabType	Back End was Fab
Status(製造中を除く)	

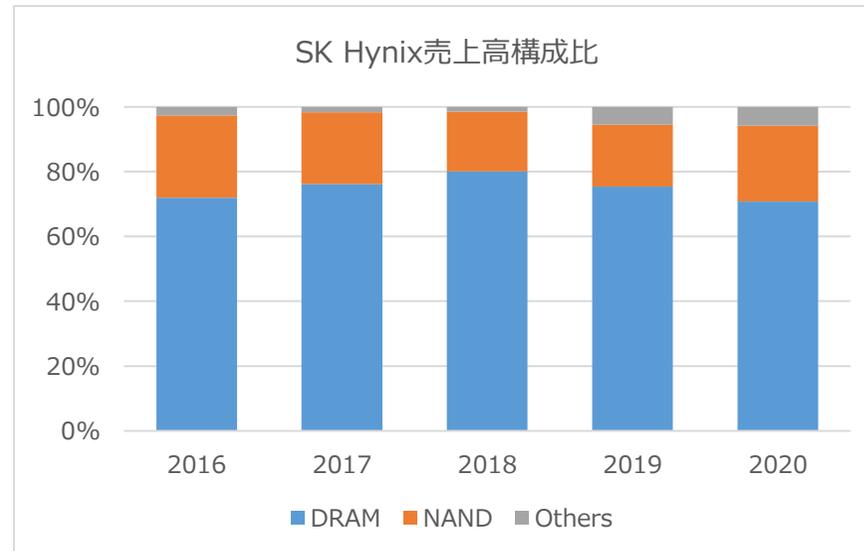
※オレンジの網掛けは後工程ファブ

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-3-4.SK Hynix (韓国)

1-2-3-4-1. SK Hynix プロファイル

- SK Hynix社は、韓国においてSamsungについて売上の高い半導体製造企業である。
- 2016年後半頃からサーバ向けのメモリを中心として需要が増加し、2018年にはSK Hynixの歴史の中で最も高い収益を達成している。2018年までがスーパーサイクルであったため、最も高い売上となっている。
- 2016年と2020年の増減率はIDM企業の中で最も高かったが、これは2016年がメモリ需要が低迷した時期であったことと、2020年のDRAM需要は最高売上である2018年より落ちているものの、モバイルNANDやゲーム機用のSSDの製品の販売が拡大したことからNANDの売上が高くなり、2016年の売上より高くなっていたことが理由だと考えられる。
- **半導体収益の7割がDRAM**である。2020年にはDDR4型と比較して1.8倍の速度、20%の消費電力削減を実現したDDR5型DRAMを発売した。
- 2020年、Intel社のNAND事業部門全体を買収することを決定した。
- 2021年、京畿道龍仁市に大規模の半導体クラス「M16」工場が完成した。**EUVプロセスを用いた1nmのDRAM製品の製造が計画**されている。
- 1か国・地域に後工程ファブを有する。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるため0を追加した。

(資料) ICInsights、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-4-2. SK Hynix Fabロケーション



1カ国・地域に工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(WFFに情報があつたもののみ記載)

※"+ "は製造開始に至っていない工場を示す。

※後工程の工場関連の情報があつたものは赤字で示す。

(資料) World Fab Forecast(SEMI)、ニュースサイトの情報をもとにNTTアドバンステクノロジーが作成

1-2-3-4-3. SK Hynix Fab詳細情報 (1/4)

【Memory】

工場名	C2	C2F (new Fab Wuxi originally C2 expansion)	M10 (DRAM) + R3 (R&D)	M11+M12 (2D NAND) CF	M11+M12 (3D NAND) CF	M14 (3D NAND)	M14 (DRAM)
所在地	Wuxi, Jiangsu, China	Wuxi, Jiangsu, China	Icheon, Gyeonggi, Korea	Cheongju, Chungcheongbuk, Korea	Cheongju, Chungcheongbuk, Korea	Icheon, Gyeonggi, Korea	Icheon, Gyeonggi, Korea
テクノロジー	CMOS	DRAM	CMOS	CMOS, Cu			
Wafer size (inch)	12	12	12	12	12	12	12
Geometry (microns)	0.021	0.017	0.018	0.014	0.03	0.03	0.016
3D Layer					128	128	
Fab Type	Fab	Fab	R&D, Fab	Fab	Fab/Split	Fab	Fab
Status (製造中を除く)							

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-4-3. SK Hynix Fab詳細情報 (2/4)

工場名	M15 (phase 1)	M15 (phase 2)	M16 (phase 1) (first floor)	M16 (phase 2)	M17 (phase 1)	M17 (phase 2)
所在地	Cheongju, Chungcheongbuk, Korea	Cheongju, Chungcheongbuk, Korea	Icheon, Gyeonggi, Korea	Icheon, Gyeonggi, Korea	Yongin, Gyeonggi, Korea	Yongin, Gyeonggi, Korea
テクノロジー	NAND		EUV	EUV	4D NAND	4D NAND
Wafer size (inch)	12	12	12	12	12	12
Geometry (microns)	0.03	0.03	0.016	0.014	0.03	0.03
3D Layer	128	256			324	352
Fab Type	Fab	Fab	Fab	Fab	Fab	Fab
Status (製造中を除く)		Planned		Planned	Planned	Planned

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-4-3. SK Hynix Fab詳細情報 (3/4)

					【Opt】	【etc】
工場名	M18 (phase 1)	M18 (phase 2)	M19 P1+P2	M20 P1+P2	M10 (CIS)	R 3
所在地	Yongin,Gyeonggi,Korea	Yongin,Gyeonggi,Korea	Yongin,Gyeonggi,Korea	Yongin,Gyeonggi,Korea	Icheon,Gyeonggi,Korea	Icheon,Gyeonggi,Korea
テクノロジー		3D	DRAM or NAND	DRAM or NAND	CMOS	CMOS, Cu
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.012	0.03	0.01	0.01	0.04	0.021
3DLayer		2XXL	2XXL	2XXL		
FabType	Fab	Fab	Fab	Fab	Fab/Split	R&Dのみ
Status(製造中を除く)	Planned	Planned	Planned	Planned		

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-4-3. SK Hynix Fab詳細情報 (4/4)

【Foundry】

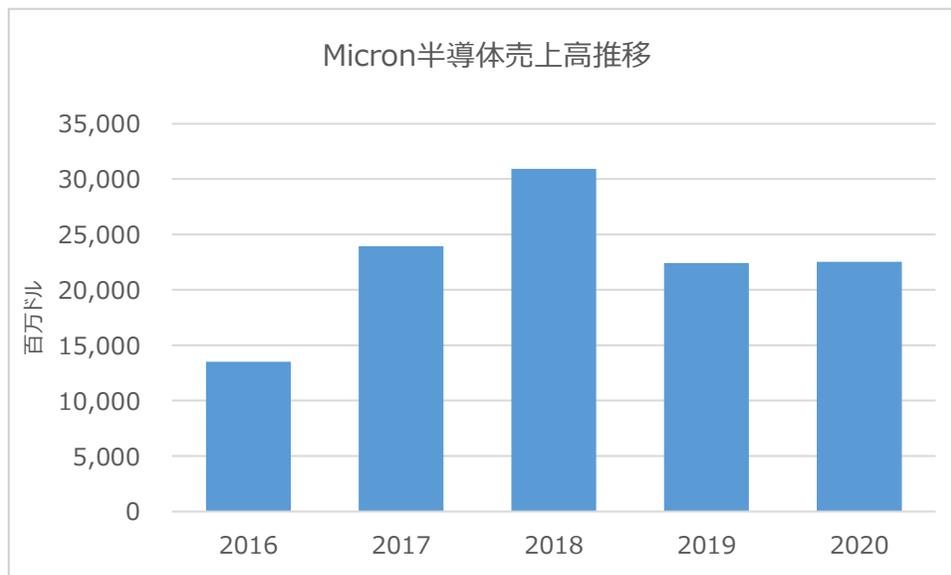
工場名	M8 (SK Hynix System IC)	Wuxi JV fab (System IC)
所在地	Cheongju, Chungcheongbuk, Korea	Wuxi, Jiangsu, China
テクノロジー	CMOS	
Wafer size(inch)	8	8
Geometry (microns)	0.09	0.09
3DLayer		
FabType	Fab	Fab
Status(製造中を除く)	Will be closed	

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-5.Micron (米国)

1-2-3-5-1. Micron プロファイル

- Micron社は米国に本社を持つDRAMメモリの製造を中心とした半導体製造企業である。
- メモリの製造を中心としているため、2018年までスーパーサイクルによる影響で売上高が増加した。
- DRAMでは現在1Znmの量産を行っており、1Znmより微細化した「1α(ワンアルファ)」と呼ばれるメモリチップを出荷することを2020年に発表した。
- NANDではRG技術を用いた128層の3D NANDが量産段階に入った。
- 2022年1月、業界初となる176層QLC NANDの量産出荷を発表した。
- 13カ国・地域に製造工場を持ち、そのうちの4カ国・地域には後工程ファブを有する。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるため0を追加した。

(資料) ICInsights、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-5-2. Micron Fabロケーション



6カ国・地域に工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(WFFに情報があつたもののみ記載)

※"+ "は製造開始に至っていない工場を示す。

※後工程のファブ関連の情報があつたものは赤字で示す。

(資料) World Fab Forecast(SEMI)、企業HPや年次報告書等をもとにNTTアドバンステクノロジーが作成

1-2-3-5-3. Micron Fab詳細情報 (1/4)

【Memory】

工場名	Building 60 (3D Xpoint)	Expansion Fab	Fab 10A (Phase 1 + 2)	Fab 10N - 3D portion	Fab 10W 3D-part (former Fab 7, Tech)	Fab 10X (Singapore)
所在地	Lehi,Utah,US	Manassas,Virginia,US	Singapore,Singapore,Singapore	Singapore,Singapore,Singapore	Singapore,Singapore,Singapore	Singapore,Singapore,Singapore
テクノロジー	CMOS, Cu, low-k, high-K	DRAM, 2D NAND	CMOS, 3D NAND	3D NAND	CMOS, Cu, low-k	3D NAND
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.02	0.02	0.02	0.02	0.02	0.02
3DLayer			128	128	96	176
FabType	Fab	Fab, Pilot, R&D	Fab, R&D	Fab, Pilot/Split	Fab/Split	Fab
Status(製造中を除く)						

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-5-3. Micron Fab詳細情報 (2/4)

工場名	Fab 11 (Inotera Fab 1)	Fab 11 (Inotera Fab 1a)	Fab 11 (Inotera Fab 2)	Fab 15 (was Elpida E300)	Fab 15 expansion B-Building	Fab 15 expansion F-Building
所在地	Taoyuan,Northern Taiwan,Taiwan	Taoyuan,Northern Taiwan,Taiwan	Taoyuan,Northern Taiwan,Taiwan	Hiroshima,Hiroshima,Japan	Hiroshima,Hiroshima,Japan	Hiroshima,Hiroshima,Japan
テクノロジー	CMOS	CMOS	CMOS	CMOS	CMOS, DRAM	CMOS, DRAM
Wafer size(inch)	12	12	12	12	12	12
Geometry (microns)	0.015	0.015	0.015	0.015	0.015	0.014
3DLayer						
FabType	Fab	Fab	Fab	Fab, R&D	Fab	Fab
Status(製造中を除く)						Equipping

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-5-3. Micron Fab詳細情報 (3/4)

工場名	Fab 16 A1 (former R1)	Fab 16 A2 (former R2)	Fab 16 A3 (phase 1 + phase 2)	Fab 16 mod 4 (A5)	Fab 2 3D part (IMFT)
所在地	Taichung, Central Taiwan, Taiwan	Lehi, Utah, US			
テクノロジー	CMOS, Cu, low-K	CMOS, Cu, low-K	CMOS, Cu, low-K		CMOS, Cu, low-k, high-K
Wafer size (inch)	12	12	12	12	12
Geometry (microns)	0.015	0.015	0.015	0.01	0.02
3D Layer					3DXP
Fab Type	Fab	Fab	Fab	Fab	Fab/Split
Status (製造中を除く)				Planned	

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-5-3. Micron Fab詳細情報 (1/4)

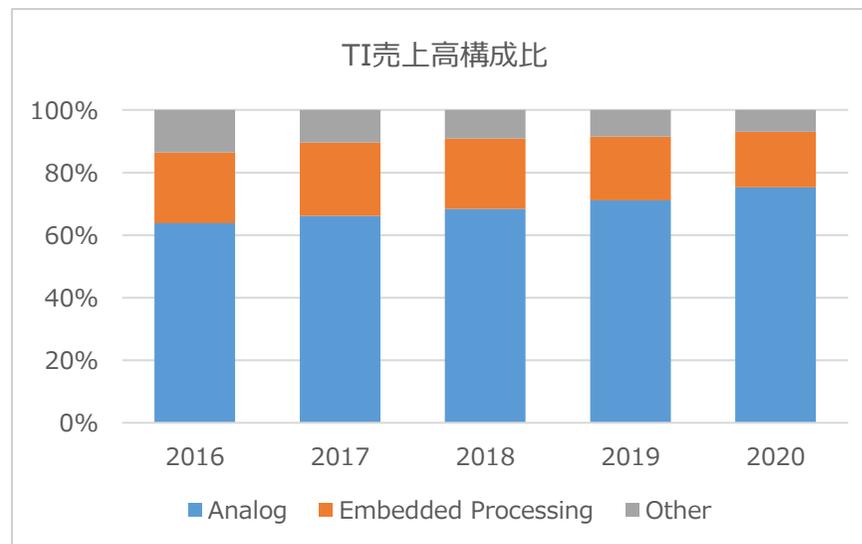
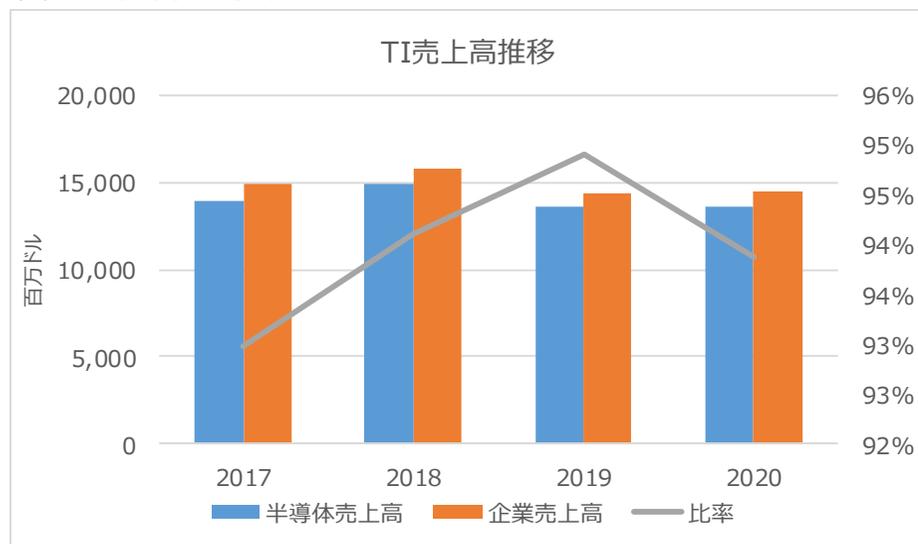
工場名	Fab 4 (Bldg 24 C,D,E)	Fab 4 (Bldg. 50)	Fab 4 (Bldg. 51 R&D expansion)	Fab 6 (MTV DRAM)	Fab 6 (MTV Flash, former IMFT)
所在地	Boise,Idaho,US	Boise,Idaho,US	Boise,Idaho,US	Manassas,Virginia,US	Manassas,Virginia,US
テクノロジー	CMOS, Cu, low-K, High-K	CMOS, Cu, low-K, High-K	CMOS, Cu, low-K, High-K	CMOS, Cu, low-K, high-K	CMOS, Cu, low-k, high-K
Wafer size(inch)	12	12	12	12	12
Geometry (microns)	0.016	0.01	0.01	0.03	0.02
3DLayer	176	176	176		
FabType	Fab, Pilot, R&D	R&D, Pilot	R&D, Pilot	Fab, R&D	Fab
Status(製造中を除く)					

(資料) World Fab Forecast(SEMI)の情報をもとにN T Tアドバンステクノロジーが作成

1-2-3-6.TI (米国)

1-2-3-6-1. TI プロファイル

- Texas Instruments(TI)は米国に本社を持つ、主にアナログチップと組込プロセッシングの製造等を行う企業である。
- 事業セグメントとしてはアナログと組込プロセッシングとその他に分けられる。アナログにはパワー半導体、シグナルチェーンや大容量製品ラインが含まれ、組込プロセッシングにはマイクロコントローラとプロセッサのラインが含まれる。
- 2020年現在産業用が売上の37%、車載用が売上の20%を占める。
- 2019年はどの事業も減少しているが、特に組込プロセッシング事業の売上が落ちている。米中貿易摩擦により需要が落ちていることが大きな要因の一つと考えられる。
- アナログについては近年力を入れており、2021年にMicronの米国にある300mmウェーハのファブを買収予定であることを、また今後米国のテキサスに300mmウェーハのファブを建設予定であることを発表している。300mmウェーハのファブでの生産は200mmファブに比べて20%のコスト削減につながるという。
- 現在世界各地に14の製造拠点をもち、そのうちの7工場には後工程が含まれる。工場所在地を次スライドに、工場の詳細情報を次々スライドに示す。



・2016年は十億単位で公表されていたため百万ドル単位に合わせるため0を追加した。

(資料) ICInsights、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-6-2. TI Fabロケーション



8カ国・地域に工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(WFFに情報があつたもののみ記載)

※“+”は製造開始に至っていない工場を示す。

※赤字は後工程ファブがある地域を示す。

(資料) World Fab Forecast(SEMI)、ICInsights、企業HPをもとにNTTアドバンステクノロジーが作成

1-2-3-6-3. TI Fab詳細情報(1/3)

	【Memory】	【Analog】			
工場名	SP1	C FAB (was SMIC Fab 11A)	D FAB (South building)	DMOS 5	DMOS 6
所在地	Aizu Wakamatsu, Fukushima, Japan	Chengdu, Sichuan, China	Dallas, Texas, US	Dallas, Texas, US	Dallas, Texas, US
テクノロジー	CMOS, Cu, Mirrorbit	CMOS, BiCMOS, PMIC, MOSFET	BICMOS; Cu, MEMS	CMOS, Cu, MEMS	CMOS, Cu, Low-K
Wafer size(inch)	12	8	8	8	12
Geometry (microns)	0.065	0.35	0.4	0.13	0.065
3DLayer					
FabType	Fab	Fab	Fab	Fab, Pilot, R&D	Fab, R&D
Status(製造中を除く)	On hold				

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-6-3. TI Fab詳細情報(2/3)

工場名	F FAB	JV3	M Fab (was National)	R Fab 1
所在地	Freising,Bavaria,Germany	Aizu Wakamatsu,Fukushima,Japan	South Portland,Maine,US	Richardson,Texas,US
テクノロジー	CMOS, BiPolar, BiCMOS, SiGe	CMOS, Cu	CMOS; BiCMOS; SiGe	CMOS, BiCMOS
Wafer size(inch)	8	8	8	12
Geometry (microns)	0.13	0.13	0.18	0.065
3DLayer				
FabType	Fab	Fab	Fab	Fab
Status(製造中を除く)				

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-3-6-3. TI Fab詳細情報(3/3)

【MEMS】

工場名	R Fab 2	S FAB	DHC (Dallas Heater Chip) Line	Miho 8 (was Miho 6)+ Miho 5
所在地	Richardson, Texas, US	Sherman, Texas, US	Dallas, Texas, US	Inashiki, Ibaraki, Japan
テクノロジー	CMOS, BiCMOS	Bipolar	MEMS	CMOS, MEMS, DLP
Wafer size (inch)	12	6	8	8
Geometry (microns)	0.065	0.8	1	0.18
3D Layer				
Fab Type	Fab	Fab	R&D, Pilot	Fab
Status (製造中を除く)	Constructing			

(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

1-2-4. Fabレス市場

1-2-4-1. Fabレス上位企業の売上規模ランキング（Top5）

- Fabレス上位5社の半導体売上(2017年-2020年)を示す。
- 売上高上位3社は年ごとにランキングの入れ替えが起こっている。
- 2020年5位であるAMDは、PC・サーバ向けCPUでシェアを伸ばし増減率が最も高い86%となっている。プロセッサRyzenが市場に受け入れられ、Intelのシェアを奪っている。
- NVIDIAは主力であるGPUがそもそものPCゲーム向け用途の他にスーパーコンピュータでの利用、暗号通貨のマイニングでの利用など広く使われシェアを伸ばしている。

Fabレス上位5社の売上高推移（単位：百万ドル）

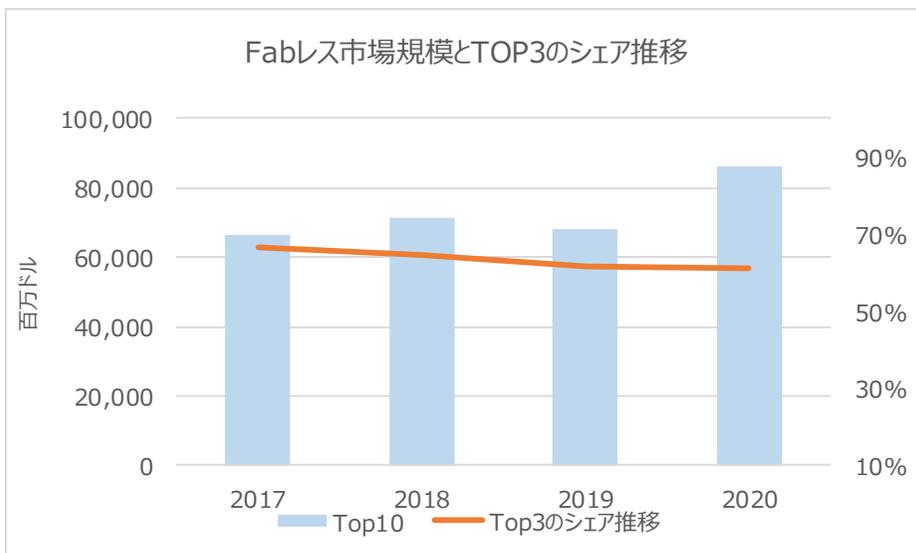
	2017	2018	2019	2020	2020年/2017年 増減率	主な製品
Qualcomm(米国)	17,029	16,370	14,518	19,407	14%	モバイルSoC
Broadcom(米国)	18,453	18,547	17,246	17,745	-4%	無線・通信インフラ向け半導体製品
NVIDIA(米国)	8,691	11,163	10,125	15,412	77%	GPU
MediaTek(台湾)	7,941	7,882	7,962	10,929	38%	モバイルSoC
AMD(米国)	5,253	6,475	6,731	9,763	86%	CPU、GPU

- 売上高を公表している企業のみ表示している
- Broadcomの収益は半導体事業のみである
- Qualcommの収益はQCT事業のみであり、QTLは含まれていない。
- NVIDIAの収益にはOEM/IP事業を、Fabレスとして計上している

（資料） TrendForceの情報をもとに N T T アドバンステクノロジーが作成

1-2-4-2. Fabレスの売上規模ランキング

- Fabレス上位10社の売上(2017年-2020年)推移及びTop3の上位10社の中でのシェア推移を示す。
- 2017年からの増減率ではRealtek、AMD、NVIDIAが大きく売上を伸ばしており、特にAMDとNVIDIAは2019年と比較しても増減率が40%を超える結果となっている。
- 下左図に示されたTop3のシェア推移は年々下がっているが、これは①2020年現在2位のBroadcomの売上があまり伸びていないこと、②Top3以外の企業が近年売上を伸ばしていることが要因である。



注) Top3 : 右表にあげた上位10社のトータル売上にしめるシェア推移 (右軸)

(資料) TrendForceの情報をもとにNTTアドバンステクノロジーが作成

Fabレス上位10社の売上高 (単位: 百万ドル)

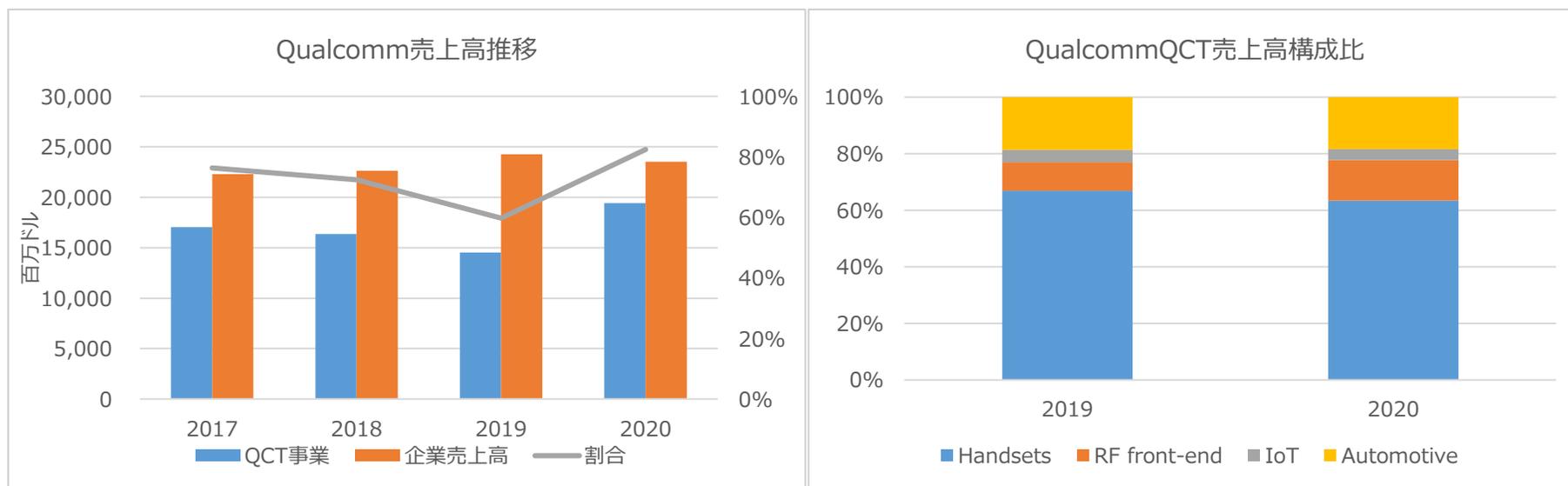
	2017	2018	2019	2020	20/17増減率
Qualcomm(米国)	17,029	16,370	14,518	19,407	14%
Broadcom(米国)	18,453	18,547	17,246	17,745	-4%
NVIDIA(米国)	8,691	11,163	10,125	15,412	77%
MediaTek(台湾)	7,941	7,882	7,962	10,929	38%
AMD(米国)	5,253	6,475	6,731	9,763	86%
Xilinx(米国)	2,438	2,868	3,234	3,053	25%
Marvel(米国)	2,392	2,823	2,708	2,942	23%
Novatek(台湾)	1,585	1,813	2,085	2,712	71%
Realtek(台湾)	1,376	1,518	1,965	2,635	91%
Dialog(英国)	1,353	1,442	1,421	1,376	2%
Top10Total	66,511	70,901	67,995	85,974	29%

- 売上高を公表している企業のみを示している
- Broadcomの収益は半導体事業のみである
- Qualcommの収益はQCT事業のみであり、QTLは含まれていない
- NVIDIAの収益にはOEM/IP事業を、Fabレス売上高として計上している

1-2-4-3. Qualcomm (米国)

1-2-4-3-1. Qualcomm プロファイル

- Qualcomm社は米国に本社を持つ、通信技術開発および半導体の設計等を行う企業である。
- 2020年においてはFabレス企業の中で売上が最も高い。
- 事業としては主に半導体事業を扱うQCT事業、ライセンスに関する事業であるQTL事業等があり、QCT事業はQualcommの売上の7割程を占める。QCT事業の中で売上が最も高く、売上に占める割合が高いのはHandsets(携帯端末向け製品)だが、RFフロントエンド関連の売上也近年増加している。
- 2020年度の売上が高かった要因の一つとしては、**2020年度後半からAppleとの複数年にわたるチップセット供給契約に基づく出荷を開始したことが考えられる。**



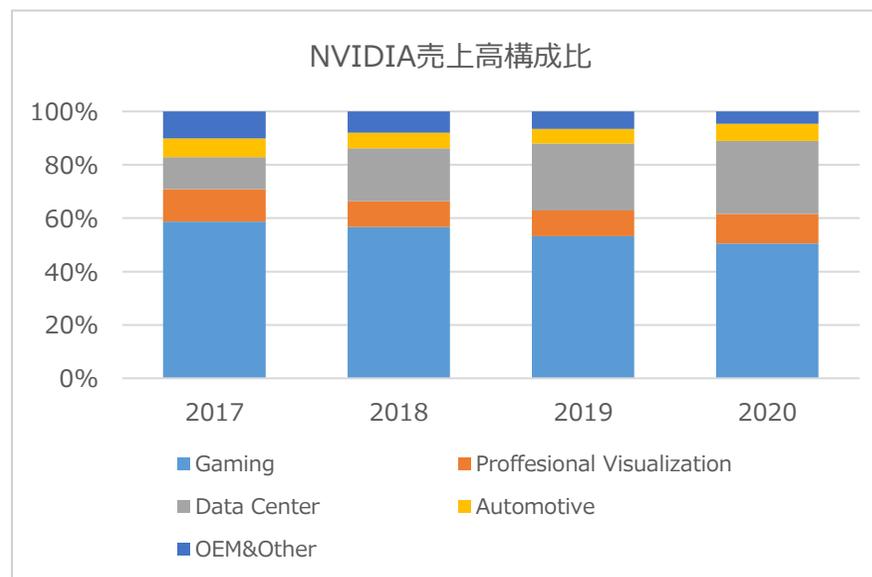
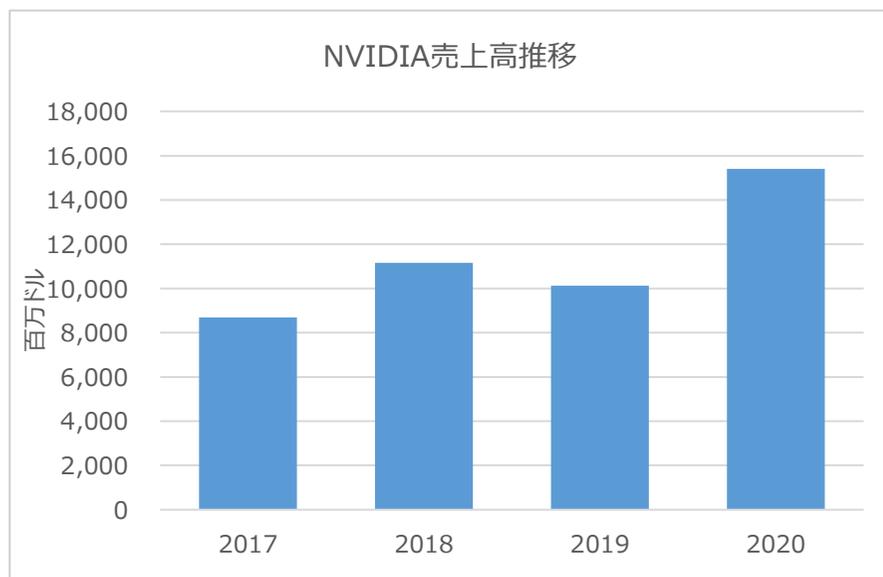
・QCT事業のみであり、QTLは含まれていない

(資料) TrendForce、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-4-4.NVIDIA (米国)

1-2-4-4-1. NVIDIA プロファイル

- 米国に本社を持つ半導体関連のファブレス企業である。
- NVIDIAはコンピュータ向けのGPUやGPGPU、ゲーム機用のグラフィックボード等を開発・販売をしている。
- 売上の割合として高いのはゲーム市場向けとデータセンター市場向けである。ゲーム市場向けは売上げのおおよそ50%を占めている。データセンター市場向けは2017年以降伸びており、企業の売上の割合としても30%近く、ゲーム市場向けに追いつきそうな勢いである。データセンター強化のためか2020年にイスラエルの半導体企業Mellanox Technologies社を買収している。
- 2020年にデータセンター向けプロセッサDPU(Data Processing Unit)を、2021年にはデータセンター向けCPUを発表している。
- AI関連に関してはIBM、Microsoftと提携しており、GPU関連ではCiscoやDell Technologiesといった大手ベンダーに採用されている。



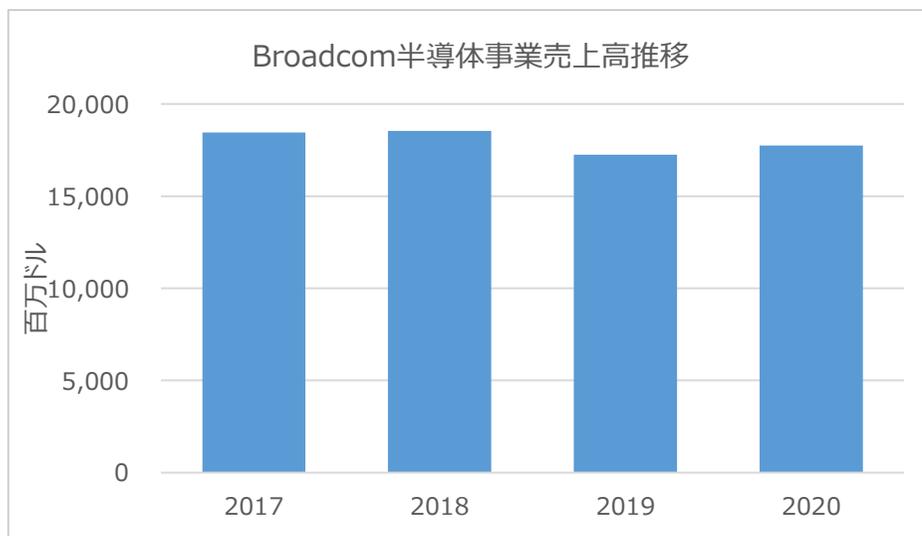
・NVIDIAの収益はOEM/IP事業を、Fabレス売上高として計上している

(資料) TrendForceの情報をもとにNTTアドバンステクノロジーが作成

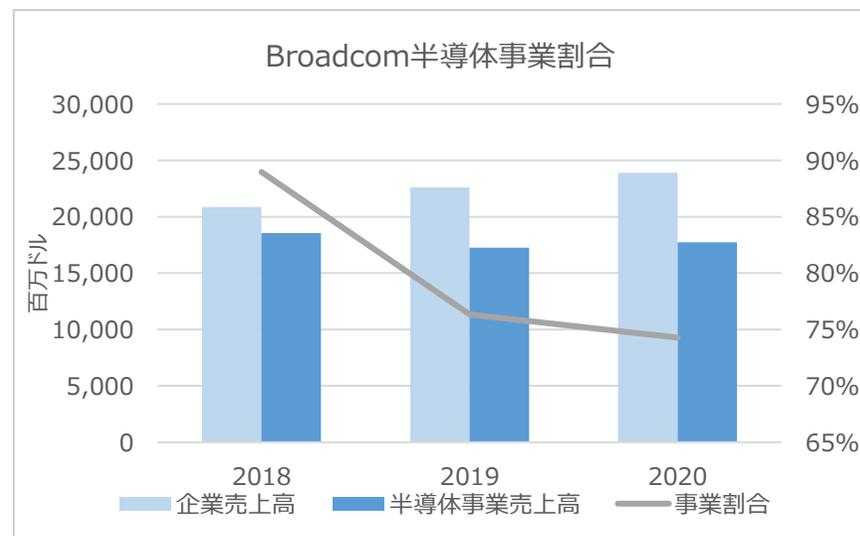
1-2-4-5. Broadcom (米国)

1-2-4-5-1. Broadcom プロファイル

- Broadcomは米国に本社を持つインフラストラクチャソフトウェアや半導体の設計開発、販売を行うファブレス企業である。
- 半導体においてはデジタルおよびミックスドシグナルのCMOSベースの製品と、III-V族ベースのアナログ製品を中心とした半導体製品を開発している。
- 2020年の半導体事業の売上は2019年よりも増えているが、企業全体から見る半導体事業の割合としては減少傾向にある。これはCOVID-19による影響と、2018年後半以降に**米中貿易摩擦が発生したことにより、Huawei社等の中国サプライヤーに販売ができなくなったことが原因**である。2020年度の地域別の純収益の情報では、中国での純収益が2018年度と比べて増減率が20%ほど減少していた。



・Broadcomの収益は半導体事業のみである

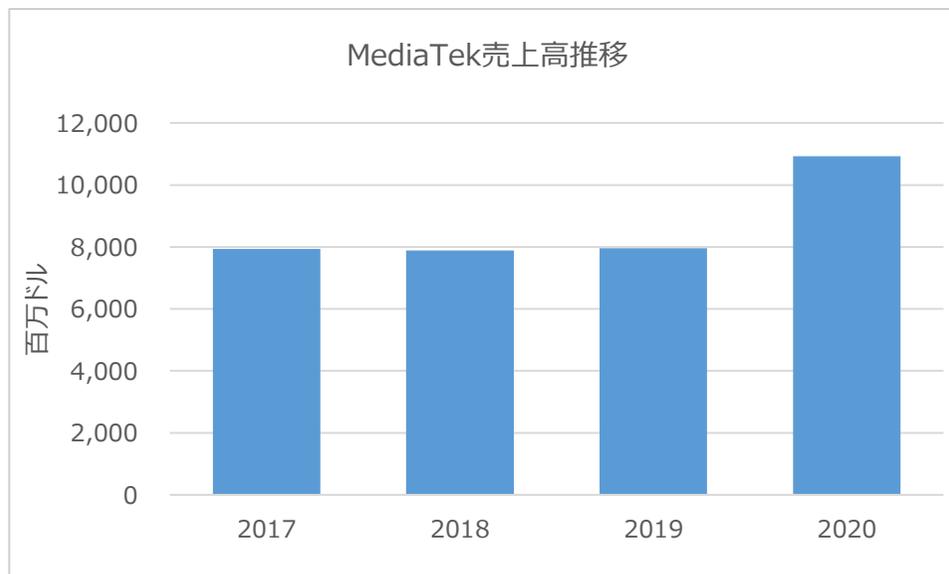


(資料) TrendForceの情報をもとにNTTアドバンステクノロジーが作成

1-2-4-6. MediaTek (台湾)

1-2-4-6-1. MediaTek プロファイル

- MediaTek社は台湾に本社を持つ主にモバイル端末やIT製品向けのSoCの設計を行っている半導体のファブレス企業である。
- 主な市場はワイヤレス通信製品市場、デジタルTV製品市場、ASIC製品市場、アナログ製品市場、ブロードバンド通信市場である。
- 2020年度の売上高は、ワイヤレス通信製品であるスマートフォン向けSoCが数多く供給されたことや、COVID-19の影響を受け、**MediaTekのチップが内蔵されたChromebook等の需要が増加したことが理由として考えられる。**
- 2020年にはIntelと共同して開発したパソコン向け5Gモデムを発表している。
- MediaTekの製品はWi-Fi6Eのテストベッドとしても選ばれており、Wi-Fi7への投資も開始しているという。

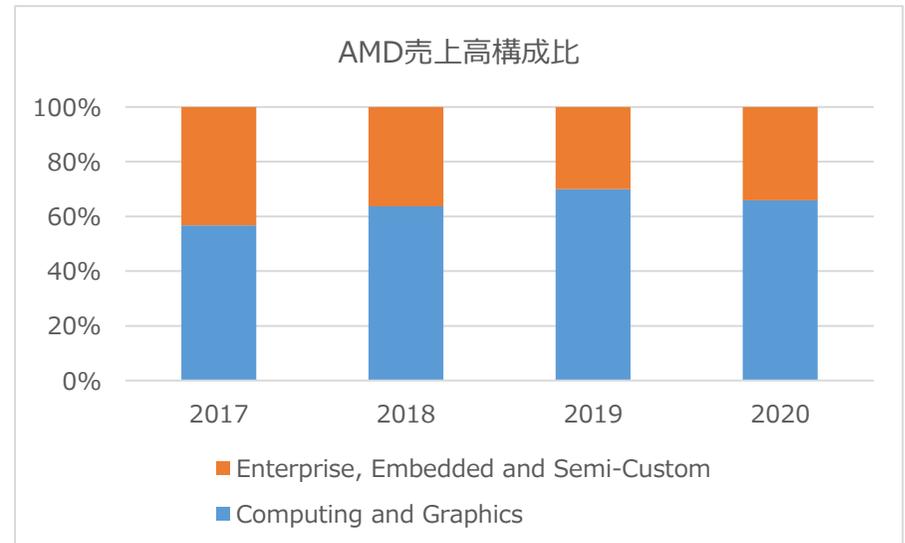
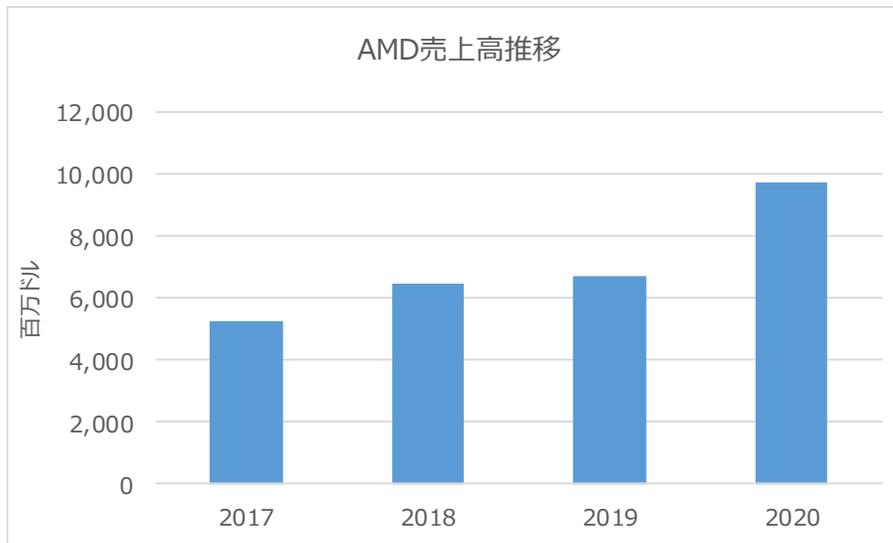


(資料) TrendForceの情報をもとにNTTアドバンステクノロジーが作成

1-2-4-7.AMD (米国)

1-2-4-7-1. AMD プロファイル

- AMD社は米国に本社を持つ主にCPUやGPUの開発を行う半導体のファブレス企業である。2008年に製造部門はGlobal foundriesに分社化した。
- x86マイクロプロセッサ、グラフィックスGPUやデータセンタ用GPUといったものやサーバーおよび組み込み用プロセッサ、セミカスタムSoC製品を提供している。
- Computing and Graphics事業については2017年と比較すると116%の伸び率となっている。これはx86プロセッサの売上が大きく貢献していることが要因だと考えられる。マイクロプロセッサ市場での主な競合はマイクロプロセッサ市場のマーケットリーダーでもあるIntelである。**2021年第4四半期にはx86プロセッサについてIntelには及ばないが25%近くのシェアを有しているという。**
- 2022年にFPGA大手のXilinxの買収が完了したことを発表した。これによりデータセンタ向けプロセッサの強化を図ると考えられる。



(資料) TrendForce、企業HP、年次報告書等の情報をもとにNTTアドバンステクノロジーが作成

1-2-5. OSAT市場

1-2-5-1. OSAT上位企業の売上規模ランキング (Top5)

- OSAT上位5社の売上(2017年-2021年(予測))を示す。
- 売上高トップのASEは5年間トップを維持し、増減率でも104%とトップになっている。M&A等で積極的にシェアを伸ばしている。
- JCETは売上高3位であるが、増減率ではTop5中最も低い。
- ランキングはこの期間変動がない。ただし、ASEに経営統合されたSPILをランキングに入れていないため、ランキングの繰り上がりがあった。
- 台湾・中国の企業が上位を占めている。

OSAT上位5社の売上高推移 (単位：百万ドル)

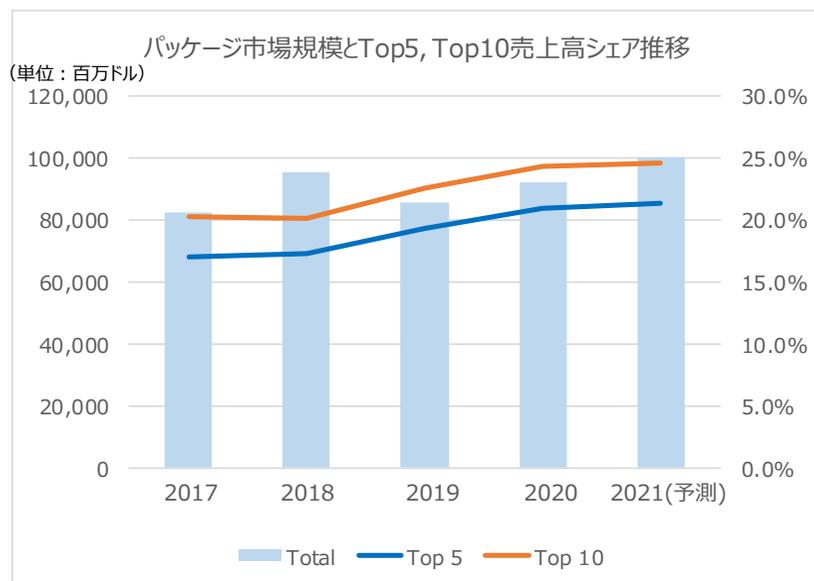
	2017	2018	2019	2020	2021 (予測)	2021年/2017年増減率
ASE (台湾) (注)	4,181	5,965	6,442	7,418	8,530	104.02%
Amkor (米国)	3,450	3,582	3,364	4,293	4,640	34.49%
JCET (中国)	3,470	3,543	3,368	3,610	3,900	12.39%
Powertech (台湾)	1,966	2,262	2,115	2,570	2,780	41.40%
Tongfu Microelectronics (中国)	956	1,088	1,165	1,400	1,510	57.95%

注) ASEは2016年に経営統合したSPIL(Siliconware Precision Industries)を含む

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-2. OSAT上位企業の売上規模ランキング (Top10)

- OSAT上位10社の売上(2017年-2021年(予測))推移およびTop5, Top10のシェア推移を示す。
- 市場全体の増減率は21.7%であるがTop10中ではJCETおよびUTAC以外は全体の伸び以上の伸びを示しており、上位企業への集中が進んでいると考えられる。
- Top10のシェアの殆どがTop5のシェアであるという昨年までと同様の傾向を示している。



注) Top10: 右表にあげた上位10社の市場全体に占めるシェア推移 (右軸)
Top 5: 右表にあげた上位5社の市場全体に占めるシェア推移 (右軸)

OSAT上位10社の売上高およびパッケージ市場推移 (単位: 百万ドル)

#		2017	2018	2019	2020	2021(予測)	21/17増減率
1	ASE (台湾) (注)	4,181	5,965	6,442	7,418	8,530	104.02%
2	Amkor (米国)	3,450	3,582	3,364	4,293	4,640	34.49%
3	JCET Group (中国)	3,470	3,543	3,368	3,610	3,900	12.39%
4	Powertech (台湾)	1,966	2,262	2,115	2,570	2,780	41.40%
5	Tongfu Microelectronics (中国)	956	1,088	1,165	1,400	1,510	57.95%
6	Tianshui Huatian Technology (TSHT, 中国)	1,013	1,054	1,135	1,150	1,270	25.37%
7	ChipMOS Technologies (台湾)	331	348	391	396	430	29.91%
8	Chipbond Technology (台湾)	407	459	530	605	660	62.16%
9	UTAC Holdings Ltd group (UTAC, シンガポール)	602	536	494	480	520	-13.62%
10	OSE (台湾)	273	283	323	370	400	46.52%
	Others	65,851	76,080	66,173	67,588	75,760	15.05%
	Total	82,500	95,200	85,500	91,900	100,400	21.70%

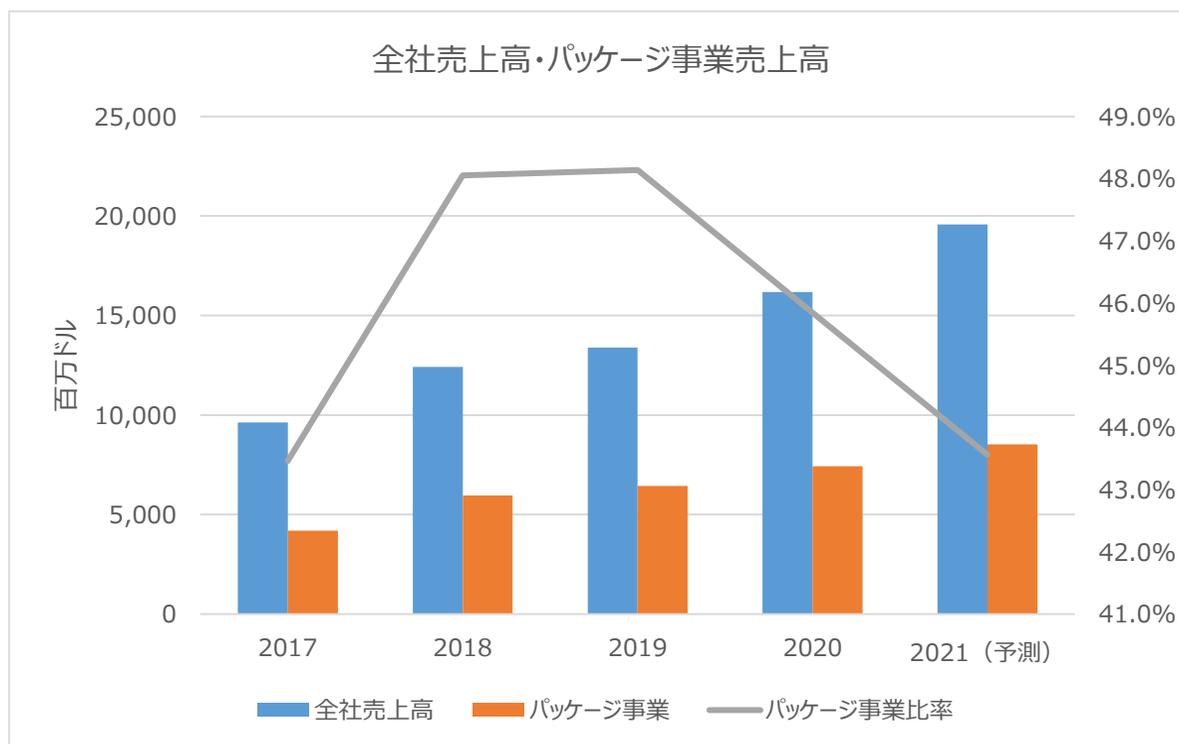
注) ASEは2016年に経営統合したSPIL(Siliconware Precision Industries)を含む

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3. ASE (台湾)

1-2-5-3-1. ASE プロファイル

- ASE社は、OSAT市場で売上高トップであるが、EMS企業USIを買収しOSAT以外の事業も行っており、パッケージ事業比率は他のOSAT上位企業に比べて低い。
- バンピングサービスやFC、WLP、SiP、FOWLP/FOPLPといった先端パッケージ分野を強化するプロジェクトを進めている。量産を行っているFOWLPとしては、eWLB、ASE社が独自開発したFOCoS(Fan Out Chip on Substrate)、M-Seriesの3種類がある。
- eWLBは、Infineon Technologies社からライセンスを受けたものである。
- 6カ国・地域に17工場を持っている。工場所在地と工場数を次スライドに、工場の詳細情報を次々スライドに示す。



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3-2. ASE Fabロケーション



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3-3. ASE Fab詳細情報 (1/4)

工場名	Kaohsing	Shanghai	ChungLi	ASEWH	ASEN
所在地	Kaohsing, Taiwan	Shanghai, China	Chung-Li, Taiwan	Shandong, China	Suzhou, China
工場動向	本社工場。2020年8月、新ラインK13建設着手、2023年竣工予定。建設費27\$Mの計画。ラインK24は研究開発機能も備える。FOWLPの生産能力は1万8,000枚/月。	SOPやQFP、QFN、BGAなど銅ワイヤボンディングによる製品化が中心。2013年に東芝子会社の子会社を買収して組み込んでいる。	成熟製品が中心。各ラインで多ピン製品に対応可能。	少ピン数・リード数の成熟製品を担当。第1棟、2棟はディスクリート、少ピン数のICに特化。第3棟は先端パッケージを製造。	NXPとの合併会社としてスタート。現在はASEの100%子会社。
対応製品	自動車、通信、コンピュータ、家電、産業、汎用・標準品	通信、コンシューマ向け製品	通信、コンピュータ、家電、汎用・標準品	パワー、自動車、通信、コンシューマ等	通信、コンシューマ、産業、汎用・標準品
ライン数	8	2	4	3	2
提供パッケージ	SIP	●		●	
	SOP/TSOP	●	●	●	
	DIP	●		●	●
	QFP	●	●	●	●
	TQFP	●	●	●	
	QFN	●	●	●	●
	COF/TCF				
	BGA/FBGA	●	●	●	●
	Stacked Die BGA	●		●	●
	Flip Chip	●		●	●
	CSP	●	●	●	
	WLCSP	●		●	●
	FOWLP/FOPLP	●			
	SiP	●			
	Power IC/Tr				
Discrete					
IGBT					
Opto					
bumps形成					

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3-3. ASE Fab詳細情報 (2/4)

工場名	ASE Kunshan	ASE Wuxi	ASE Korea	ASEジャパン	ASE Malaysia
所在地	Kunshan, China	Wuxi, China	Cyeonggi-do, Korea	Yamagata, Japan	Penang, Malaysia
工場動向	第1棟は成熟した少ピン数製品中心、第2棟は先端パッケージの製造を行っている。	2013年に東芝子会社から買収。QFP、LQFP、SOP、PDIPに特化している。	第1棟、2棟ともにRFや自動車向け製品に注力している。	2004年にNECエレクトロニクスの高畠工場を買収。今後は小型QFNの製造に特化した事業展開を図る。	QFP、TQFP、BGA、SOIC、SOJ、PDIPの組み立て及びテストを担当。
対応製品	通信、コンシューマ、コンピュータ、産業、汎用・標準品	通信、自動車、コンシューマ、オーディオ	RF、自動車	通信、コンシューマ、産業、標準品	通信、コンシューマ、産業
ライン数	2	1	2	1	1
提供パッケージ	SIP				
	SOP/TSOP	●	●	●	
	DIP	●	●	●	●
	QFP	●	●	●	●
	TQFP				●
	QFN	●		●	●
	COF/TCP				
	BGA/FBGA	●		●	●
	Stacked Die BGA	●		●	●
	Flip Chip	●		●	
	CSP	●		●	
	WLCSP	●		●	
	FOWLP/FOPLP				
	SiP				
	Power IC/Tr				
	Discrete				
	IGBT				
Opto					
パンパ形成					

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3-3. ASE Fab詳細情報 (3/4)

工場名	ASE Singapore	Da Fong Facility	Chung Shan Facility	Zhong Ke Facility	Hsinchu Facility	
所在地	Woodlands Loop, Singapore	Taichung, Taiwan	Taichung, Taiwan	Taichung, Taiwan	Hsinchu, Taiwan	
工場動向	FCBGA, HSBGA, PoP, aQFN、QFN、TQFPのテスト工程を担当。	QFP、QFNなどの成熟製品からWLCSPなどの先端パッケージまで担当。	組立、テスト、バンブ形成に対応。組立工程ではFC加工も担う。	2015年にバンブ形成工場として稼働し、その後WLCSPやFOWLP、2.5D/3Dといった先端パッケージの製造も行っている。	テスト工程を担当。	
対応製品	通信、コンシューマ、産業	通信、自動車、コンシューマ、産業、標準品等	通信、自動車、コンシューマ、産業、標準品等	通信、自動車、コンシューマ、産業、標準品等	通信、自動車、コンシューマ、産業、標準品等	
ライン数	1	2	2	1	2	
提供パッケージ	SIP					
	SOP/TSOP					
	DIP					
	QFP	●(TEST)	●			●(TEST)
	TQFP	●(TEST)	●			●(TEST)
	QFN	●(TEST)	●			●(TEST)
	COF/TCP					
	BGA/FBGA	●(TEST)	●	●	●	●(TEST)
	Stacked Die BGA	●(TEST)	●	●	●	●(TEST)
	Flip Chip		●	●	●	●(TEST)
	CSP		●	●	●	●(TEST)
	WLCSP		●	●	●	●(TEST)
	FOWLP/FOPLP				●	●(TEST)
	SIP				●	
	Power IC/Tr					
	Discrete					
	IGBT					
Opto						
バンブ形成			●	●		

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-3-3. ASE Fab詳細情報 (4/4)

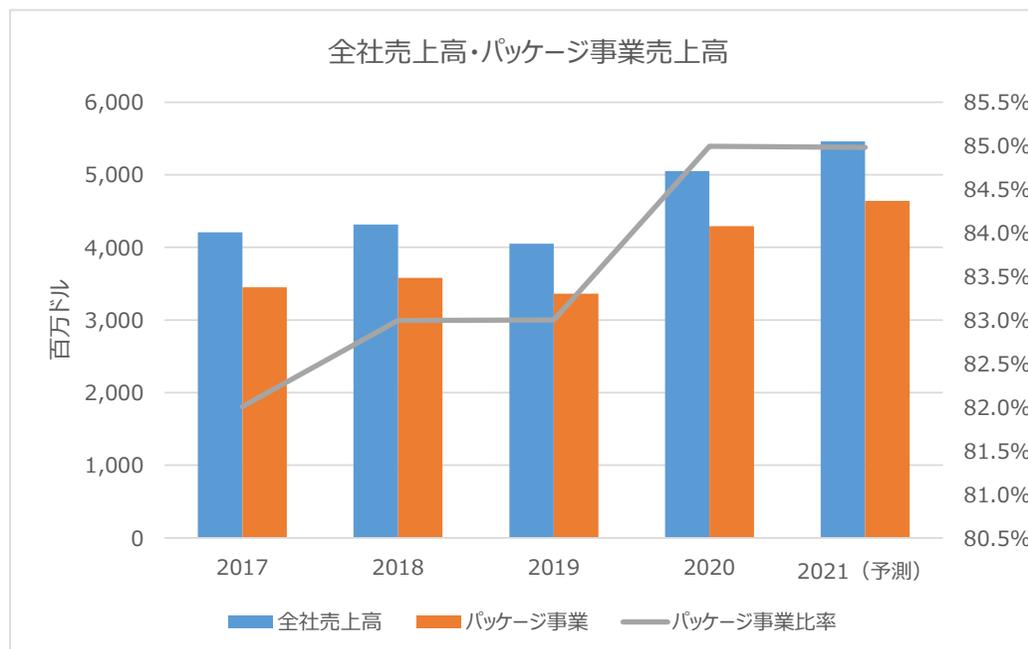
工場名	Changhua Facility	SuZhou Facility
所在地	Changhua, Taiwan	SuZhou, China
工場動向	先端パッケージの製造拠点。	紫光集団が資本参加している。第1棟、2棟はワイヤボンディングによる成熟製品向け加工が中心。第3棟はFC、WLCSP、WLBGAなどの先端パッケージの製造拠点となっている。
対応製品	ロジック	通信、産業、家電等
ライン数	2	3
提供パッケージ	SIP	●
	SOP/TSOP	●
	DIP	●
	QFP	●
	TQFP	●
	QFN	●
	COF/TCP	
	BGA/FBGA	●
	Stacked Die BGA	●
	Flip Chip	●
	CSP	●
	WLCSP	●
	FOWLP/FOPLP	
	SiP	●
	Power IC/Tr	
	Discrete	
IGBT		
Opto		
バンブ形成		

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-4. Amkor (米国)

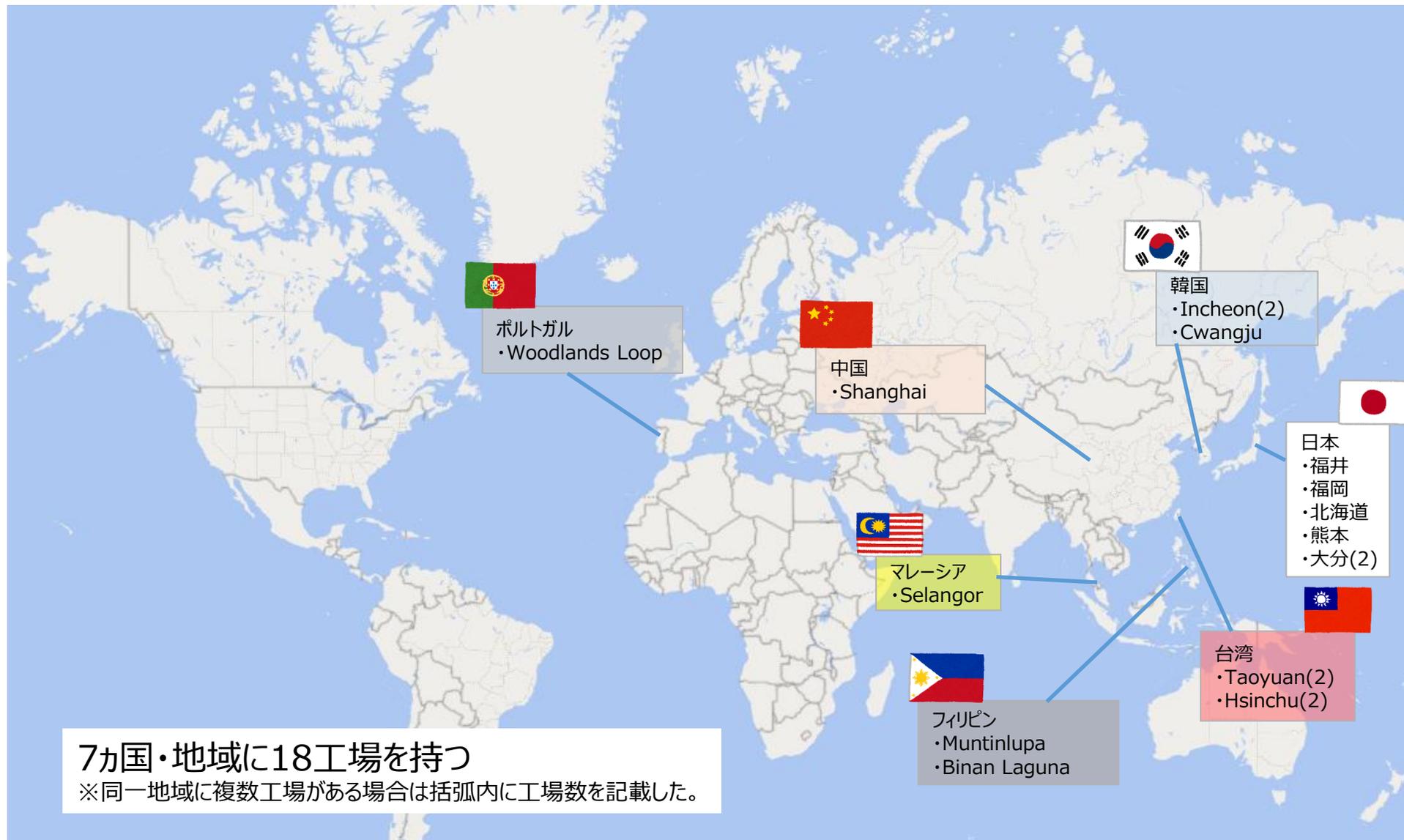
1-2-5-4-1. Amkor プロファイル

- Amkor社の対応業務は組立プロセス、テストプロセス、モジュール形成にまで及んでおり、前工程を終えた後の後工程をワンストップで提供できる。積極的な買収戦略による業容拡大を図っている。
- 主な顧客としては、Intel、ルネサスエレクトロニクス、STMicro、TI、キオクシア、Broadcom、Qualcomm、Samsung、TSMCなどがある。上位顧客で売上高の63%を占める。
- FOWLの独自技術としてSWIFTを開発している。
- ファインピッチでは、従来のはんだバンプに代わって銅ピラーの採用を進めており、OSAT企業の中でも最も積極的に開発・導入している。
- 特殊用途やMEMS向けなどの多様なパッケージの対応も進めており、FusionQuadパッケージを2018年に開発している。
- 7カ国・地域に18工場を持っている。工場所在地と工場数を次スライドに、工場の詳細情報を次々スライドに示す。



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-4-2. Amkor Fabロケーション



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-4-3. Amkor Fab詳細情報 (1/3)

工場名	K3	K4	K5	P1	P3/P4	C3(上海工場)	
所在地	Incheon, Korea	Gwangju, Korea	Incheon, Korea	Muntinlupa, Philippines	Binan Laguna, Philippines	Shanghai, China	
工場動向	テスト工程専業。	フリップチップ(FC)、ラミネート、リードフレーム、SiP、WLPなど成熟製品および先端製品の組立、ウエハテストおよびファイナルテストを担当。	FC、WLPなどの先端線品の製造拠点。	リードフレームタイプの成熟製品を担当。	テスト工程専業。	中国市場向けを中心に組立、テストを担当。テスト工程ではQualcomm Technology社との合弁ラインを設けている。	
対応製品	ロジック、アナログ	ロジック、メモリ	ロジック、ミックスドシグナル、アナログ	ロジック、ミックスドシグナル、アナログ	ロジック、ミックスドシグナル、アナログ	ロジック、メモリ	
ライン数	1	1	1	1	1	1	
提供パッケージ	SIP	●(TEST)	●	●	●	●(TEST)	●
	SOP/TSOP	●(TEST)	●		●	●(TEST)	●
	DIP	●(TEST)	●		●	●(TEST)	●
	QFP	●(TEST)	●		●	●(TEST)	●
	TQFP	●(TEST)	●		●	●(TEST)	●
	QFN	●(TEST)	●				●
	COF/TCP						
	BGA/FBGA	●(TEST)	●	●	●	●(TEST)	●
	Stacked Die BGA	●(TEST)	●	●			●
	Flip Chip	●(TEST)	●	●			●
	CSP	●(TEST)	●	●			●
	WLCSP	●(TEST)	●	●			●
	FOWLP/FOPLP	●(TEST)	●				
	SiP	●(TEST)	●	●			●
	Power IC/Tr						
	Discrete						
IGBT							
Opto							
ハンブ形成			●				

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-4-3. Amkor Fab詳細情報 (2/3)

工場名	T1	T3	T5	T6	M1	E1	福井工場
所在地	Taoyuan, Taiwan	Hsinchu, Taiwan	Hsinchu, Taiwan	Taoyuan, Taiwan	Selangor, Malaysia	Vila do Conde, Portugal	福井, Japan
工場動向	FC、WLPなどの先端パッケージの製造拠点。バンブ形成も行う。	WLCSPの製造拠点。バンブ形成も行う。	WLCSPの製造拠点。バンブ形成も行う。	テスト工程およびダイシングプロセスを担当。	2013年の東芝エレクトロニクス・マレーシアの買収で取得。	2017年のポルトガルNanium社買収により取得。WLCSP、FOWLPの製造を担当。	パワーディスクリートの製造拠点
対応製品	ロジック、メモリ	ロジック、メモリ	ロジック、メモリ	ロジック、メモリ	アナログ、ディスクリート	ロジック	パワーディスクリート
ライン数	1	1	1	1	1	1	1
提供パッケージ	SIP				●		
	SOP/TSOP				●		
	DIP				●		
	QFP				●		
	TQFP				●		
	QFN						
	COF/TCP						
	BGA/FBGA						
	Stacked Die BGA						
	Flip Chip	●			●(TEST)		
	CSP	●			●(TEST)		
	WLCSP	●	●	●	●(TEST)		●
	FOWLP/FOPLP						●
	SIP				●(TEST)		
	Power IC/Tr						●
Discrete						●	
IGBT							
Opto							
バンブ形成	●	●	●				

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-4-3. Amkor Fab詳細情報 (3/3)

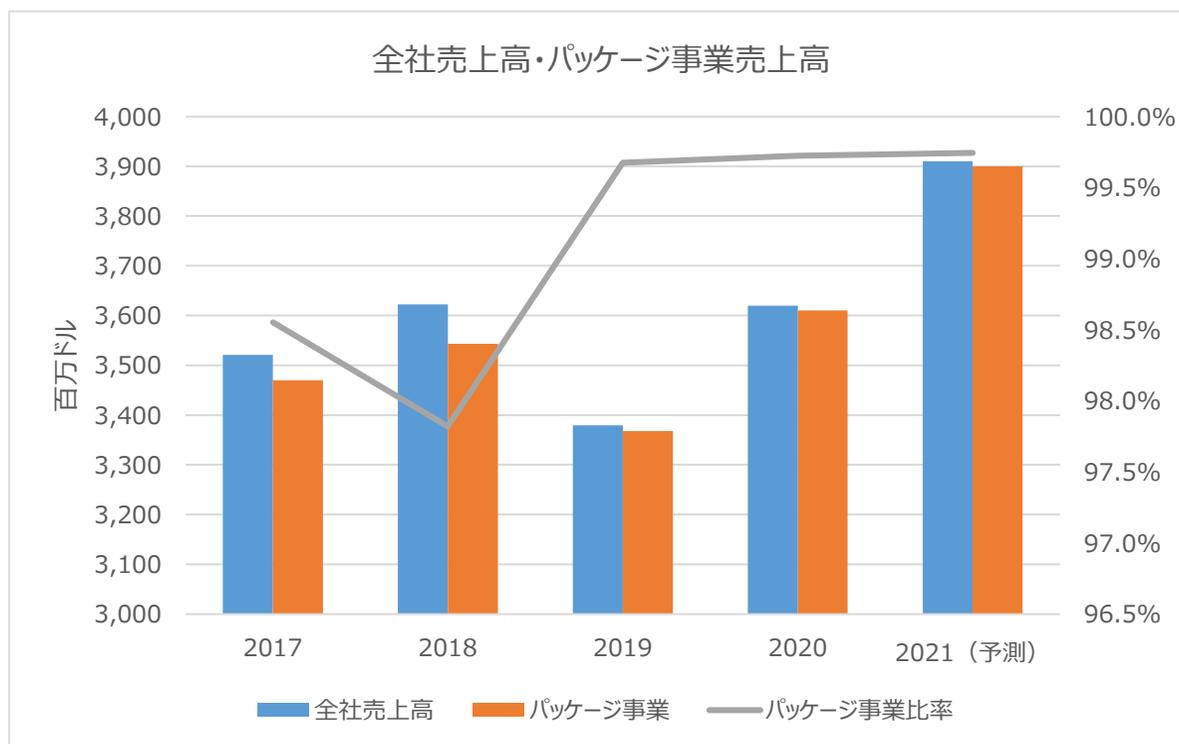
工場名	福岡工場	北海道工場	熊本工場	大分工場	臼杵工場
所在地	福岡, Japan	北海道, Japan	熊本, Japan	大分, Japan	大分, Japan
工場動向	リードフレームタイプの成熟製品を担当。	リードフレームタイプ、ラミネートタイプの成熟製品を担当。	ルネサスエレクトロニクスの熊本工場で製造しているマイコン、ロジックの製造が中心。	東芝セミコンダクター & ストレージの大分工場内に設けられている。ウエハテストを担当。	旧ジェイデバイスの本社工場。リードフレームタイプの成熟製品に加えてCMOSイメージセンサの組立も行う。
対応製品	ロジック	ロジック、アナログ	マイコン、ロジック	ロジック、アナログ	ロジック、アナログ、CMOSイメージセンサ
ライン数	1	1	1	1	1
提供パッケージ	SIP	●	●	●	●
	SOP/TSOP	●	●	●	●
	DIP	●	●	●	●
	QFP	●	●	●	●
	TQFP	●	●	●	
	QFN	●	●	●	
	COF/TCP				
	BGA/FBGA	●	●	●	
	Stacked Die BGA				
	Flip Chip			●	●
	CSP				●
	WLCSP				●
	FOWLP/FOPLP				
	SiP			●	
	Power IC/Tr				
	Discrete				
	IGBT				
	Opto				●
バンプ形成					

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-5. JCET (中国)

1-2-5-5-1. JCET プロファイル

- JCETはディスクリートの組立事業からスタートした企業で、その後IC処理を事業に組み込んだ。近年はFCBGA、WLCSP、3D/TSVといった先端パッケージ、イメージセンサなどの高付加価値製品サービスの比重を高めており、ディスクリート、低価格ICなどの従来製品から事業のシフトを進めている。
- 成熟製品が中心であったが、シンガポールSTATS ChipPACの買収が先端パッケージへのシフトの契機となった。先端パッケージでは、WLCSPやバンプ形成(300mmウェーハ)、SiP、FC、FOWLPに注力しており、生産能力が拡大している。
- 3カ国・地域に6工場を持っている。工場所在地と工場数を次スライドに、工場の詳細情報を次々スライドに示す。



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-5-2. JCET Fabロケーション



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-5-3. JCET Fab詳細情報

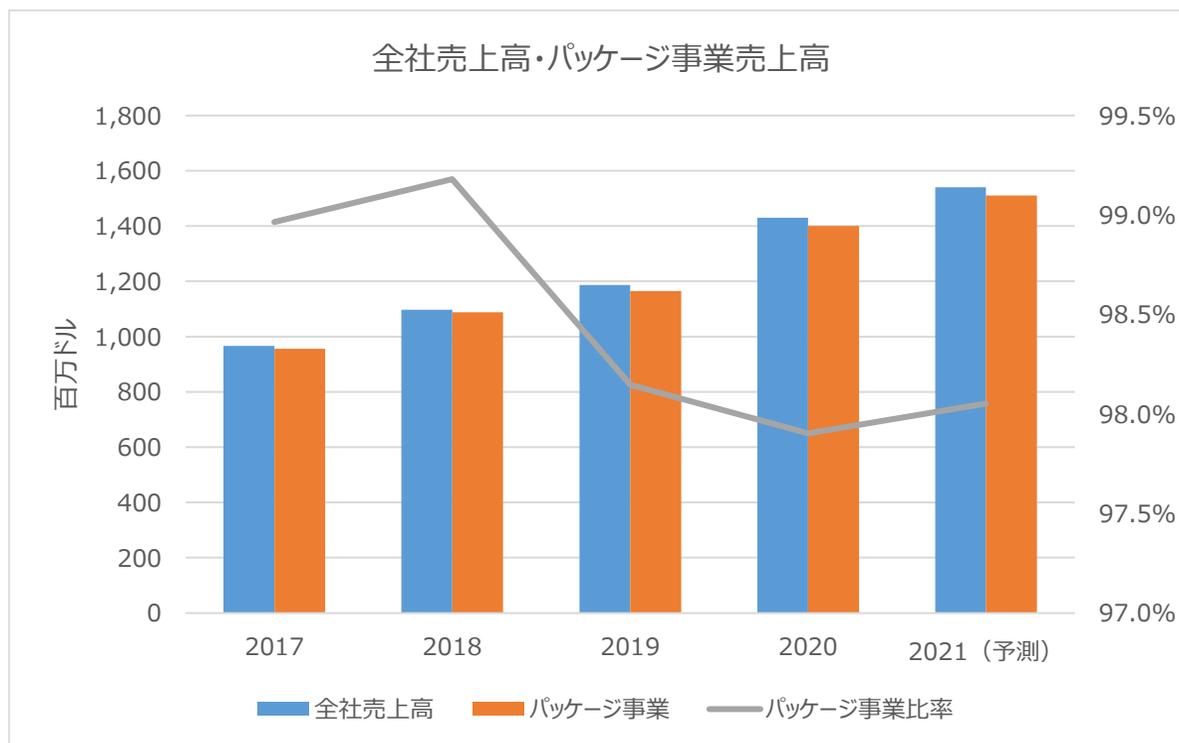
工場名	D1	D3	D8	D9	SCS(STATSChipPAC Singapore)	JCET STATS ChipPAC Korea
所在地	Jiangsu, China	Jiangsu, China	Jiangsu, China	Anhui, China	Yishun, Singapore	Incheon, Korea
工場動向	JCET, JCAPの生産拠点。	JCET, JCAP, JSCC, SJSEMIの生産拠点	2019年に既存工場の隣接地に新工場を立ち上げた。	小信号トランジスタなどのディスクリートの製造を担当。	WLCSPやFOWLP等の製造を担当。FOWLPはeWLB技術を用いて量産。	SCKおよびJSCKの生産拠点。
対応製品	ロジック	ロジック	パワーデバイス	ディスクリート	ロジック	ロジック、アナログ、MEMS等
ライン数	2	4	2	1	1	4
提供パッケージ	SIP	●				
	SOP/TSOP	●				
	DIP	●				
	QFP	●				
	TQFP	●				
	QFN	●				
	COF/TCP					
	BGA/FBGA	●				●
	Stacked Die BGA	●				●
	Flip Chip	●			●	●
	CSP	●				●
	WLCSP	●	●		●	●
	FOWLP/FOPLP	●			●	●
	SiP		●			
	Power IC/Tr			●		
	Discrete				●	
	IGBT			●		
Opto						
バンブ形成	●	●			●	●

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-6. Powertech (台湾)

1-2-5-6-1. Powertech プロファイル

- Powertech社は、元々メモリの後工程処理に特化していた企業で、特に汎用DRAMの組立事業で大きな実績を残している。メモリメーカーがNANDフラッシュの製造を強化してからは、同製品に対するサービスも強化しているほか、2011年以降はモバイル向けのロジックICやSiP、CMOSイメージセンサなどの非メモリ分野への取り組みも進めている。
- Samsung以外の手元メモリ企業は、後工程のアウトソーシングを加速しており、PowertechにもDRAMやNANDフラッシュ、双方を組み合わせたモジュールなどの製造委託が集中している。
- 4カ国・地域に15工場を持っている。工場所在地と工場数を次スライドに、工場の詳細情報を次々スライドに示す。



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-6-2. Powertech Fabロケーション



4カ国・地域に15工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-6-3. Powertech Fab詳細情報 (1/3)

工場名	Plant 1	Plant 2	Plant 3(A/C/D)	Plant 8(HSIP Plant 1)	Plant 9	Plant 10
所在地	Hsinchu, Taiwan	Hsinchu, Taiwan	Hsinchu, Taiwan	Hsinchu, Taiwan	Hsinchu, Taiwan	Hsinchu, Taiwan
工場動向	リードフレームベースの成熟製品が主体。	メモリ向け成熟製品が中心だが、B棟はサブストレートタイプの先端パッケージに注力。	Plant 2と同一敷地に新工場として建設。	台Macrotech社を買収して取得。成熟製品のパッケージが主体。	FC, WSPなどの先端パッケージの製造を担当。	2020年竣工。FOPLP専用工場。
対応製品	ロジック、アナログ	ロジック、アナログ	ロジック、メモリ、CMOSイメージセンサ	ロジック、メモリ	ロジック、メモリ	ロジック、メモリ
ライン数	1	2	3	1	1	1
提供パッケージ	SIP	●	●	●		
	SOP/TSOP	●	●			
	DIP	●				
	QFP	●	●		●	
	TQFP	●	●		●	
	QFN	●	●		●	
	COF/TCP					
	BGA/FBGA	●	●	●	●	
	Stacked Die BGA	●	●	●		
	Flip Chip	●	●	●		●
	CSP			●		●
	WLCSP		●	●		●
	FOWLP/FOPLP					●
	SiP		●	●		●
	Power IC/Tr					
	Discrete					
	IGBT					
Opto			●			
バンブ形成		●	●		●	

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにN T Tアドバンステクノロジーが作成

1-2-5-6-3. Powertech Fab詳細情報 (2/3)

工場名	Plant 11(HSIP Plant 2)	Zhunán	Toufen	Xi'an	Suzhou
所在地	Hsinchu, Taiwan	Miaoli, Taiwan	Miaoli, Taiwan	Shaanxi, China	Suzhou, China
工場動向	台Macrotech社を買収して取得。先端パッケージ製造主体。	関連会社の台Greatek Electronicsが運営。3-256ピンのリードフレームパッケージングとテスト専門。	関連会社の台Greatek Electronicsが運営。3-256ピンのリードフレームパッケージングとテスト専門。	主にMicron Technologyのメモリパッケージングを担当。	成熟製品のパッケージングを担当。
対応製品	ロジック、メモリ	ロジック、アナログ、ミックスシグナル	ロジック、アナログ、ミックスシグナル	メモリ	ロジック、メモリ
ライン数	1	2	1	1	1
提供パッケージ	SIP			●	●
	SOP/TSOP	●	●	●	●
	DIP	●	●	●	●
	QFP	●	●	●	●
	TQFP	●	●	●	●
	QFN	●			●
	COF/TCP				
	BGA/FBGA	●			●
	Stacked Die BGA	●			
	Flip Chip	●			
	CSP	●			
	WLCSP	●			
	FOWLP/FOPLP				
	SiP	●			
	Power IC/Tr				
	Discrete				
	IGBT				
	Opto				
バンブ形成					

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにN T Tアドバンステクノロジーが作成

1-2-5-6-3. Powertech Fab詳細情報 (3/3)

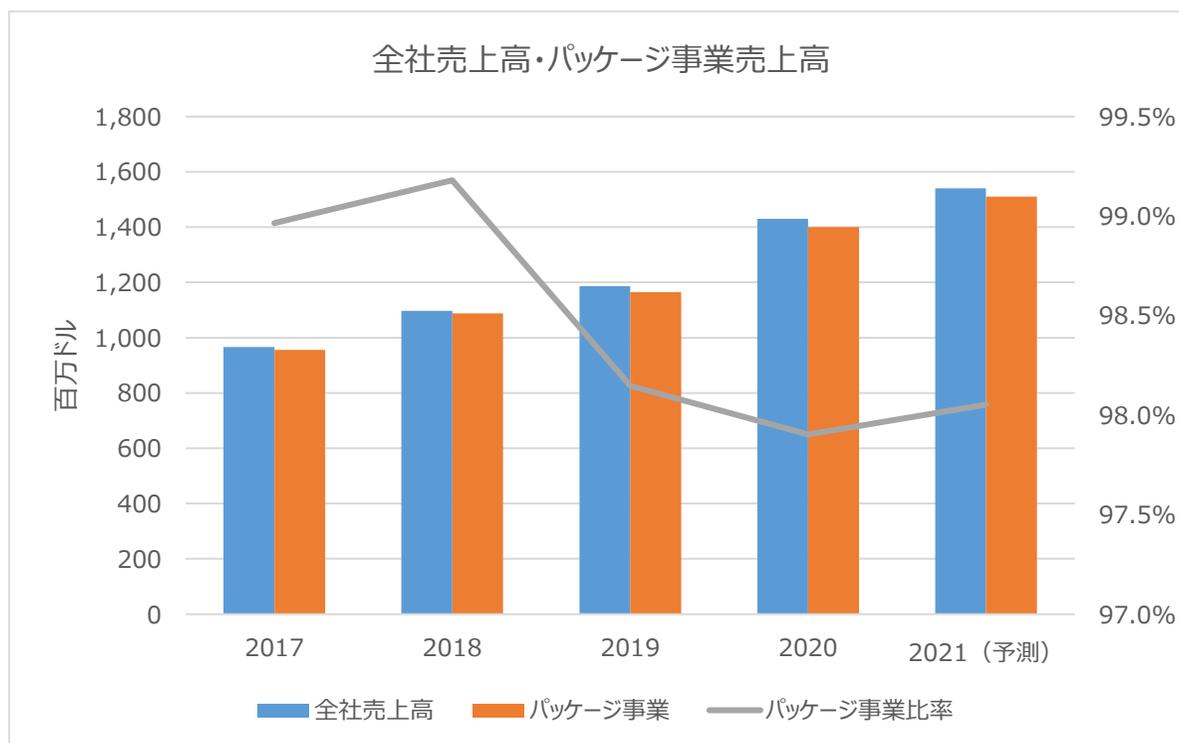
工場名	Singapore	九州事業所	会津	TeraPower
所在地	Ang Mo Kio, Shingapore	熊本, Japan	福島, Japan	Hsinchu, Taiwan
工場動向	韓Nepes社を買収して取得。FC, WLPの製造を行っており、新たにFOPLPの生産ラインを導入する計画。	2017年にテラプローブ社を子会社化。	2017年にテラプローブ社を子会社化。ウエハテスト、ファイナルテストを担当。	テラプローブの台湾子会社。ウエハテスト、ファイナルテストを担当。
対応製品	ロジック、メモリ	ロジック、メモリ、アナログ、CMOSイメージセンサ	ロジック、メモリ、アナログ、CMOSイメージセンサ	ロジック、メモリ、アナログ、CMOSイメージセンサ
ライン数	1	1	1	1
提供パッケージ	SIP	●(TEST)	●(TEST)	●(TEST)
	SOP/TSOP	●(TEST)	●(TEST)	●(TEST)
	DIP	●(TEST)	●(TEST)	●(TEST)
	QFP	●(TEST)	●(TEST)	●(TEST)
	TQFP	●(TEST)	●(TEST)	●(TEST)
	QFN	●(TEST)	●(TEST)	●(TEST)
	COF/TCP			
	BGA/FBGA		●(TEST)	●(TEST)
	Stacked Die BGA		●(TEST)	●(TEST)
	Flip Chip	●	●(TEST)	●(TEST)
	CSP	●	●(TEST)	●(TEST)
	WLCSP	●	●(TEST)	●(TEST)
	FOWLP/FOPLP	●	●(TEST)	●(TEST)
	SiP			
	Power IC/Tr			
	Discrete			
	IGBT			
Opto		●(TEST)	●(TEST)	
バンブ形成				

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-7. Tongfu Microelectronics (中国)

1-2-5-7-1. Tongfu Microelectronics プロファイル

- Tongfu Microelectronics(通富微电子)は、元々はNanton Huada Microelectronics Group社と富士通の合併企業であったが、2018年に富士通が合併から撤退し、Nanton Huada Microelectronics Groupの子会社となった。
- 2016年4月、AMDと合併会社TF-AMD(通富超威半導体有限公司)を設立している。
- AMDの7nmプロセスを用いたハイエンドプロセッサなどの売上が好調で、組立およびテストを担っている同社の売上も増加している。
- 先端パッケージの開発に注力しており、中国以外での顧客開拓を積極的に進めている。
- 2カ国・地域に6工場を持っている。工場所在地と工場数を次スライドに、工場の詳細情報を次々スライドに示す。



(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-7-2. Tongfu Microelectronics Fabロケーション



2カ国・地域に6工場を持つ

※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-5-7-3. Tongfu Microelectronics Fab詳細情報

工場名	Chongchuan	Hefei	Nantong	Xiamen	Suzhou	Penang
所在地	Jiangsu, China	Anhui, China	Jiangsu, China	Fujian, China	Suzhou, China	Penang, Malaysia
工場動向	本社工場。3工場稼働。	組立およびテスト工程を担当。	FOWLPの生産能力を確立。2019年に4Gパワーアンプの量産を開始。新たにTFBGAの量産に対応。	2019年12月、試作開始。バンブ形成、WLCSP、CP、FC、SiPの製造を計画。	FCBGA、FCPGA、FCLGA、MCMなどのハイエンドパッケージの製造を担当。	2016年にBGA、WLCSPラインの増強に着手。
対応製品	ロジック、アナログ	メモリ、ロジック、アナログ、ディスクリート	モバイル、通信機器等	ロジック、アナログ、ディスクリート	モバイル、通信機器等	ロジック、アナログ、ディスクリート
ライン数	3	2	2	1	2	2
提供パッケージ	SIP	●	●			
	SOP/TSOP	●	●			
	DIP	●	●			
	QFP	●	●			
	TQFP	●				
	QFN	●	●	●		●
	COF/TCP		●			
	BGA/FBGA	●	●	●		●
	Stacked Die BGA	●	●			●
	Flip Chip	●	●		●	●
	CSP	●	●	●	●	●
	WLCSP	●	●	●	●	●
	FOWLP/FOPLP			●		
	SiP	●	●		●	●
	Power IC/Tr					
Discrete						
IGBT						
Opto						
バンブ形成	●	●		●		

(資料) 半導体パッケージビジネス戦略2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-6. EMS市場

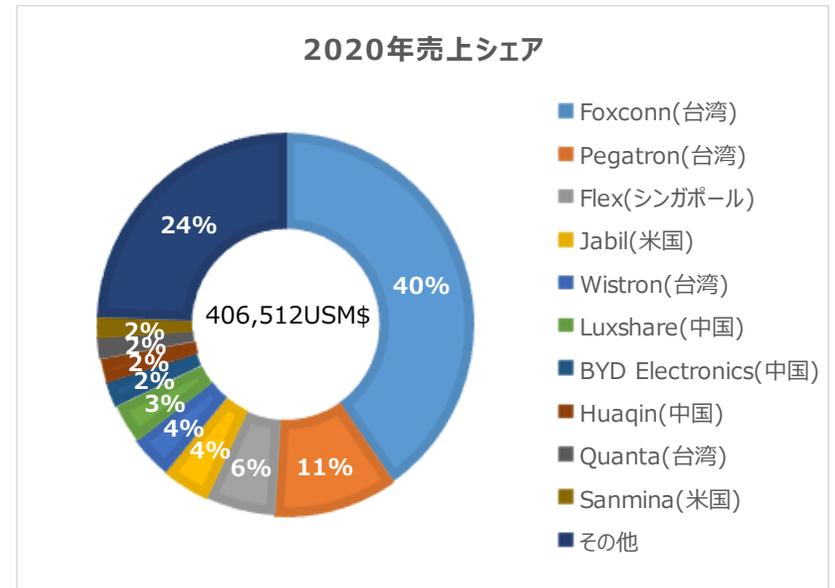
1-2-6-1. EMS上位企業の売上規模ランキング

- EMS上位10社の売上(2016年-2020年)を示す。
- 売上高トップのFoxconnは2位のPegatronに大差をつけて5年間トップを維持している。シェア40%はトップの地位が安定しているといえる。
- 2016年には10位だったApple AirPodsの組立企業であるLuxshareが2020年までに6位に飛躍的に売上高を伸ばしている。2016年から2020年までの増減率は572%となっている。2020年7月にはiPhoneの組立企業であるWistronの中国工場買収を発表^{*1}しており、2021年にはさらに順位をあげると見られる。欧米からの受託で売上を伸ばし、技術力も高めていると思われる。

(*1) <https://36kr.jp/86402/>

#	企業名	2016	2017	2018	2019	2020	20/16増減率
1	Foxconn(台湾)	133,355	144,004	161,966	163,465	163,955	23%
2	Pegatron(台湾)	35,805	34,162	38,251	40,366	43,289	21%
3	Flex(シンガポール)	24,419	24,781	25,440	26,211	24,210	-1%
4	Jabil(米国)	11,012	11,057	12,269	15,431	16,612	51%
5	Wistron(台湾)	11,403	14,447	15,371	15,176	14,602	28%
6	Luxshare(中国)	2,036	3,377	5,303	9,248	13,684	572%
7	BYD Electronics(中国)	5,783	5,987	6,247	7,897	8,882	54%
8	Huaqin(中国)	3,120	3,598	4,556	5,325	8,136	161%
9	Quanta(台湾)	6,236	6,116	6,211	6,394	7,092	14%
10	Sanmina(米国)	6,481	6,868	7,110	8,234	6,960	7%

単位：USM\$

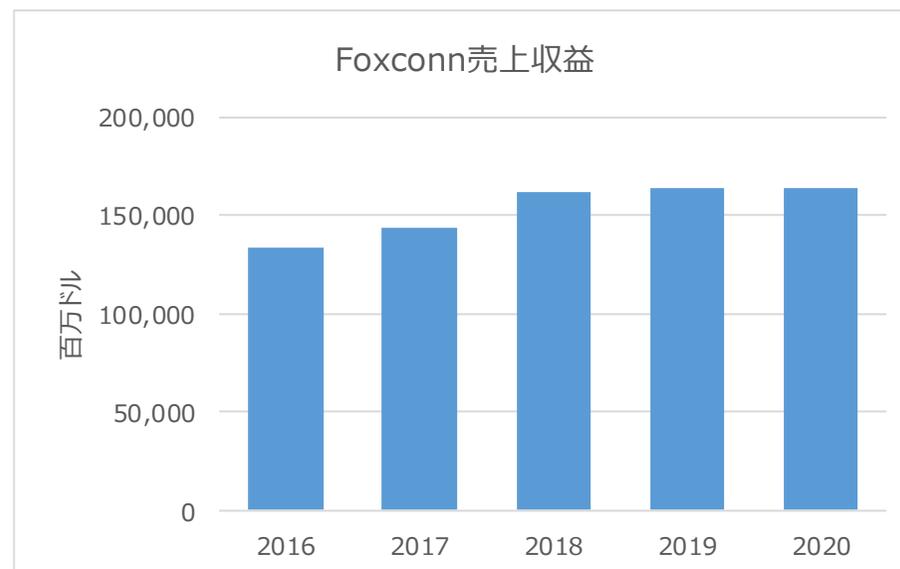


(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, Status AND FORECAST 2021-2027(QYRESEARCH社)をもとにNTTアドバンステクノロジーが作成

1-2-6-2. Foxconn (台湾)

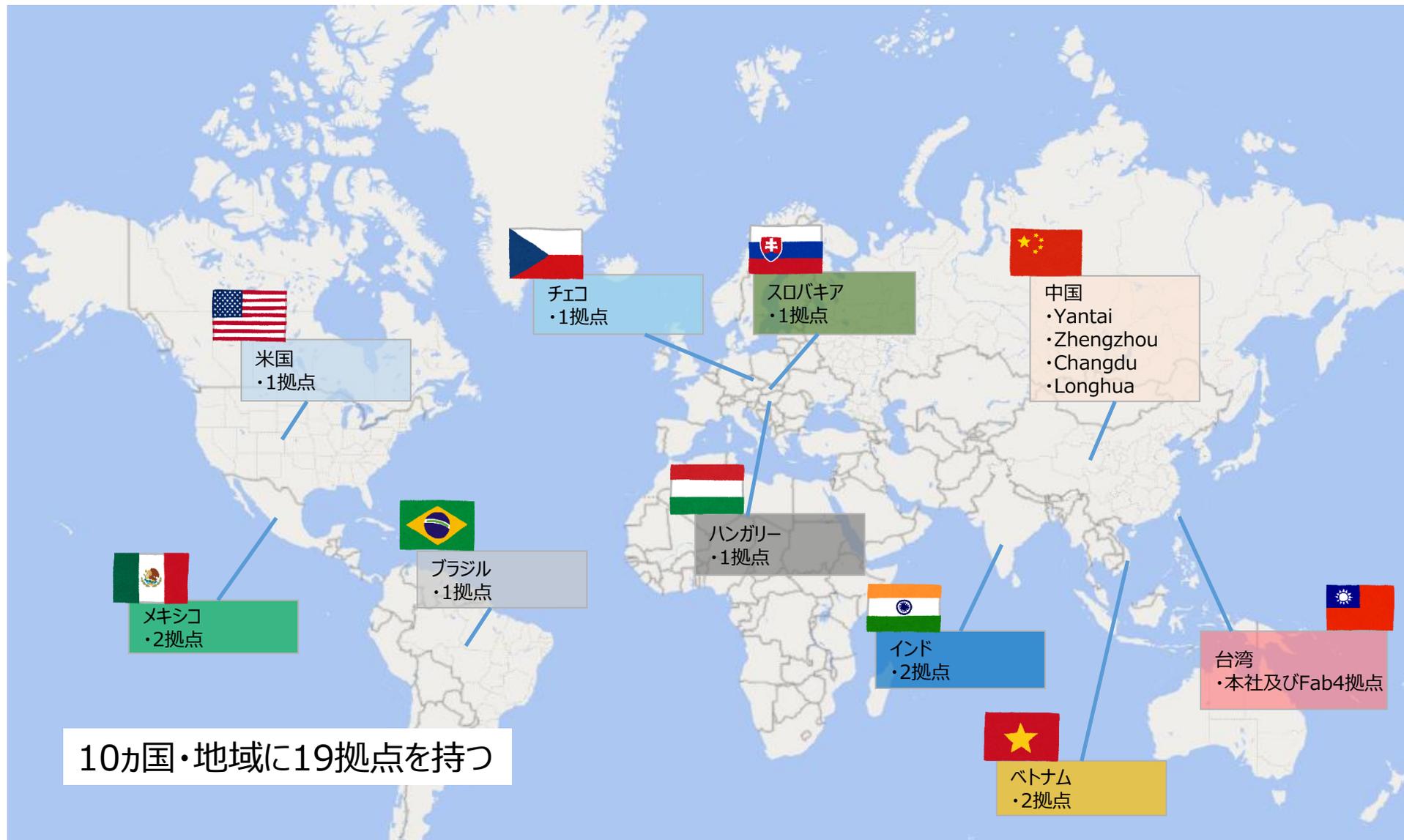
1-2-6-2-1. Foxconn プロファイル

- Foxconn(鴻海科技集団)は台湾に本社を置くEMS企業。2020年のシェアは40%となっており、2位に大差をつけてシェアトップとなっている。
- 主要工場は、台湾に6工場(虎躍工場、民生工場、頂埔一工場、頂埔二工場、頂埔五工場、南崁工場)を構える。
- Fisker(米国)、吉利汽車(中国)、BYTON(中国)等と協業して将来のEV産業での海外進出を開拓している。
- 2020年6月17日、鴻海研究院を設立し、資通安全、人工知能、量子計算、半導体、新世代通信の5大研究所を設置している。
- 2020年の研究開発費は3201百万ドル(純利益の1.76%)である。
- 4事業区分(コンシューマエレクトロニクス分野、クラウド・ネットワーク製品分野、パソコン端末分野、デバイスおよびその他の製品分野)で事業を行っている。※事業区分ごとの売上情報は非公開



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)および Foxconn2020年年次報告書をもとにNTTアドバンステクノロジーが作成

1-2-6-2-2. Foxconn ロケーション

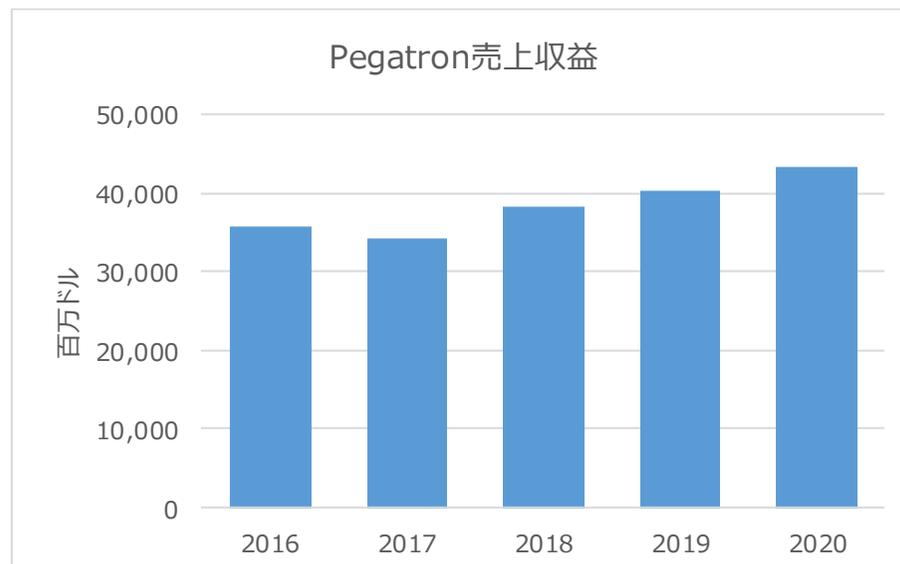


(資料) Foxconn HPをもとにNTTアドバンステクノロジーが作成

1-2-6-3. Pegatron (台湾)

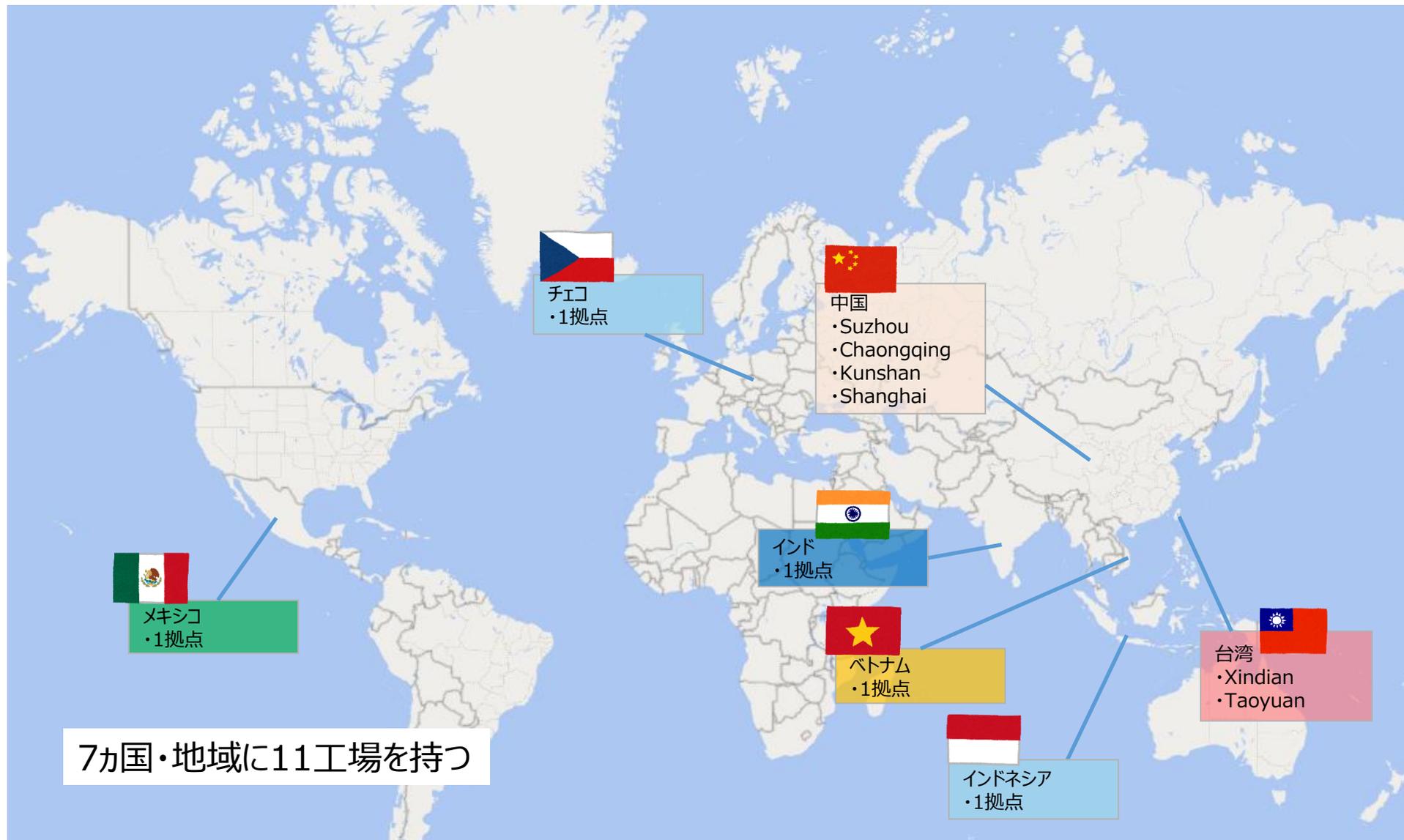
1-2-6-3-1. Pegatron プロファイル

- Pegatron(和碩聯合科技子会社)は台湾に本社を置くEMS企業である。
- 2020年の研究開発費は567百万ドル(純利益の1.19%)である。
- 地域別売上は、台湾内6.8%、アジア7.7%、欧州43.1%、米国37.6%、その他4.8%と欧州が主要市場となっている。
- 主要製品には、ノートパソコン、ネットブックコンピュータ、デスクトップコンピュータ、ゲームコンソール、ハンドヘルドデバイス、マザーボード、ビデオカード、LCD TV、およびスマートフォン、セットトップボックス、ケーブルモデムなどのブロードバンド通信製品がある。
- 情報電子製品、通信電子製品、家電等の3区分で事業を行っている。※事業区分ごとの売上情報は非公開



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)をおよび Pegatron 2020年年次報告書をもとにNTTアドバンステクノロジーが作成

1-2-6-3-2. Pegatron Fabロケーション

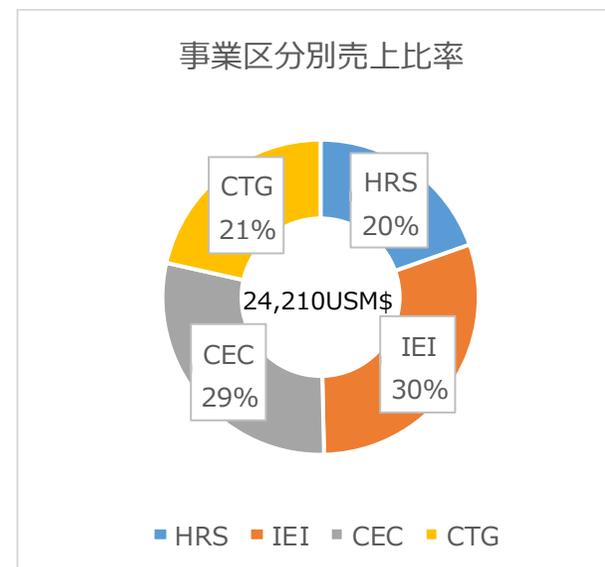
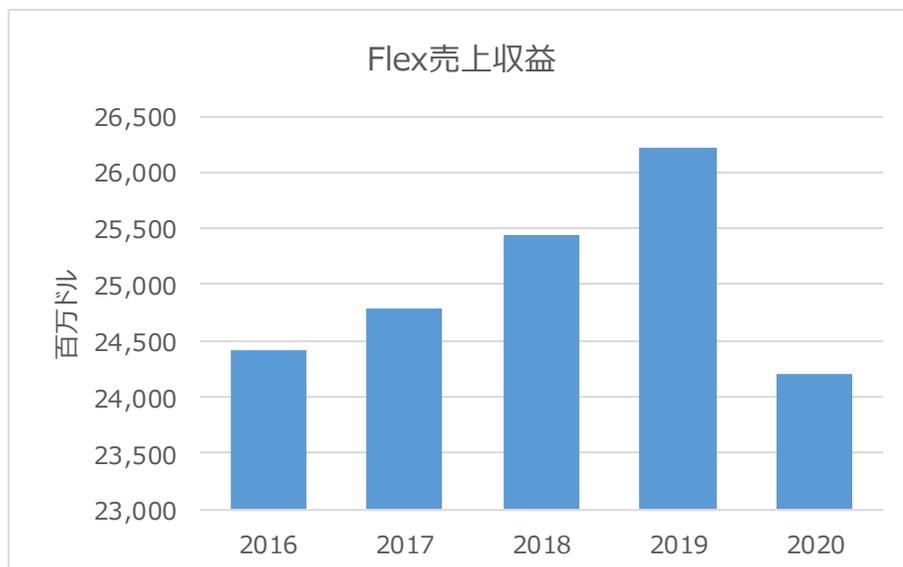


(資料) Pegatron HPをもとにNTTアドバンステクノロジーが作成

1-2-6-4. Flex (シンガポール)

1-2-6-4-1. Flex プロファイル

- Flexはシンガポールに本社を置くEMS企業である。
- 主要顧客には、Teradyne、Dyson、Xerox、Cisco、Nokia、Ericsson、Lenovo、Motorola、HP、Bose、Ford、Nexeer、Fiat Chrysler等がいる。
- HRS(High Reliability Solutions)、IEI(Industrial and Emerging Industries)、CEC(Communications & Enterprise Compute)、CTG(Consumer Technologies Group)の事業区分で事業を行っている。



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)および Flex 2020年年度報告書をもとに N T T アドバンステクノロジーが作成

1-2-6-4-2. Flex Fabロケーション



11カ国・地域に21工場を持つ

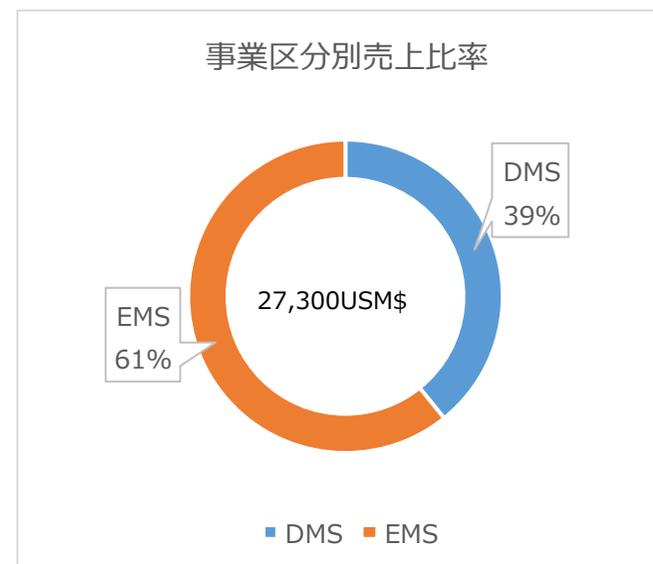
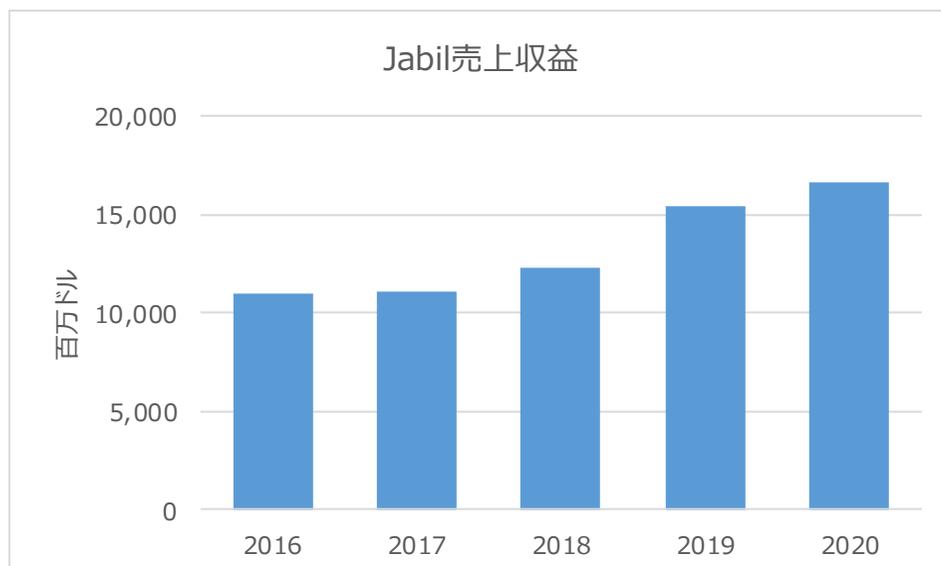
※同一地域に複数工場がある場合は括弧内に工場数を記載した。

(資料) Flex HPをもとにNTTアドバンステクノロジーが作成

1-2-6-5. Jabil (米国)

1-2-6-5-1. Jabil プロファイル

- Jabilは米国に本社を置くEMS企業である。
- Jabilは23カ国・地域に90の施設を持っている。主な生産拠点は、米国、中国、マレーシア。
- 多角化事業(DMS)と電子製造事業(EMS)の事業区分別で事業を行っている。EMS事業がJabilの柱であり、5Gおよびクラウドコンピューティング、エネルギー、産業機器等の製品を製造している。
- 主要顧客には、Amazon.com、Apple、Cisco、Hewlett-Packard、Ingenico Group、Johnson and Johnson、LM Ericsson Telephone Company、NetApp、SolarEdge Technologies、Teslaがいる。
- 2020年の研究開発費は44.1USM\$で同年の純利益の0.2%をあてている。



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)および Jabil 2020年年度報告書をもとにNTTアドバンステクノロジーが作成

1-2-6-5-2. Jabil Fabロケーション

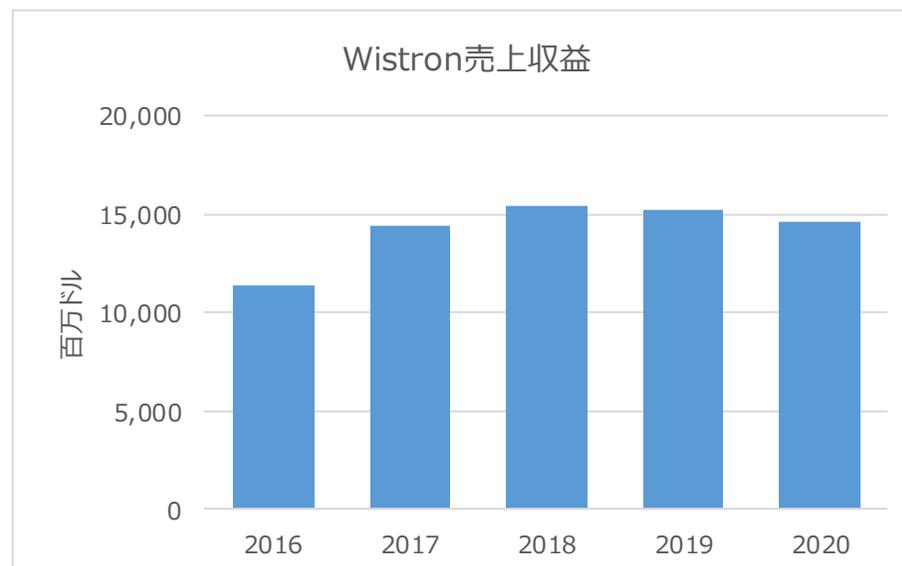


(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)をもとに
NTTアドバンステクノロジーが作成

1-2-6-6. Wistron (台湾)

1-2-6-6-1. Wistron プロファイル

- Wistronは台湾に本社を置くEMS企業である。
- 研究開発拠点は、台湾(台北、高雄)、中国 (Zhongshan、Kunshan、Shanghai、Chengdu、Chongqing、Wuhan)、米国(サンノゼ)に設置されている。
- 主な製品・サービスは、ノートパソコン、スマートフォンやハンドヘルド機器、デスクトップコンピュータとAll-in-Oneコンピュータ、ディスプレイ製品、VoIP電話機、サーバとネットワークストレージ設備、産業用PC、アフターサービス、その他。※製品種別ごとの売上情報は非公開
- 2020年の研究開発費は、647,934百万ドルで、同年の利益の2.25%となっている。



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)および Wistron 2020年年度報告書をもとにNTTアドバンステクノロジーが作成

1-2-6-6-2. Wistron Fabロケーション

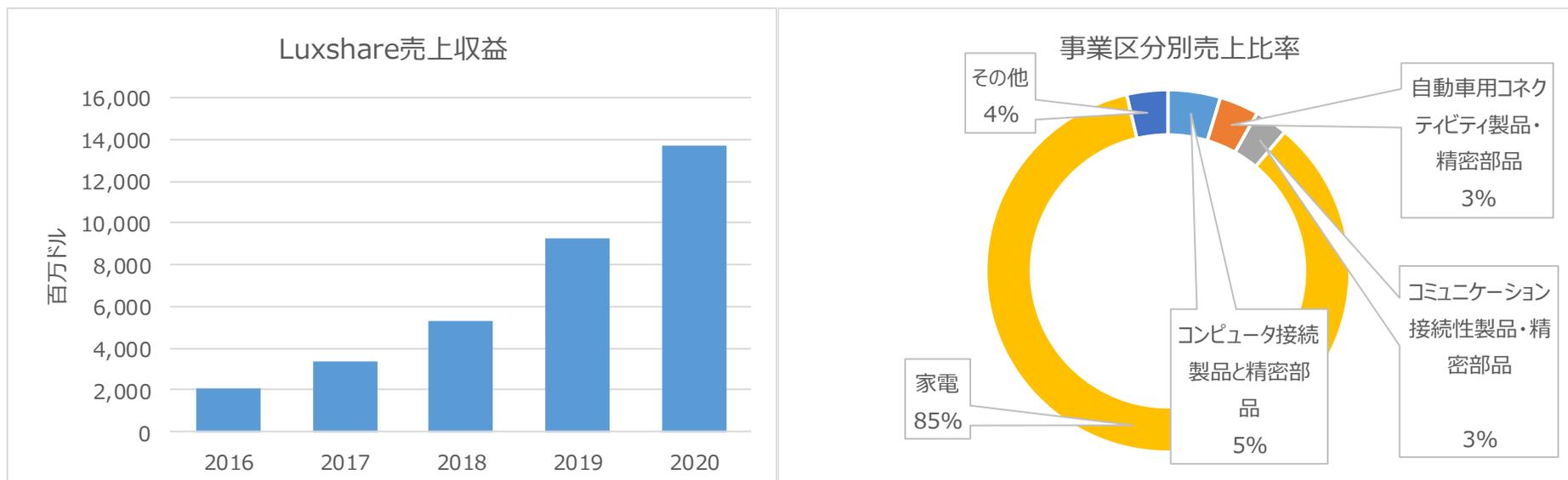


(資料) Wistron HPをもとにNTTアドバンステクノロジーが作成

1-2-6-7. Luxshare（中国）

1-2-6-7-1. Luxshare プロファイル

- Luxshareは中国に本社を置くEMS企業である。
- 東莞、昆山、台湾、米国に高度な生産技術と製造プロセスの研究開発プラットフォームを構築している。
- 2020年の研究開発費は、372787百万ドルであり、前年比62.74%増えている。
- 事業区分は、コンピュータ接続製品と精密部品、自動車用コネクティビティ製品・精密部品、コミュニケーション接続性部品・精密部品、家電、その他の区分となっている。
- 海外市場での売上が90.63%、中国国内が9.37%と海外主体の売上構成になっている。
- Wistronの中国工場を買収し、iPhone組立企業の3番手となっている。



(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)およびLuxshare 2020年年度報告書をもとにNTTアドバンステクノロジーが作成

1-2-6-7-2. Luxshare ロケーション



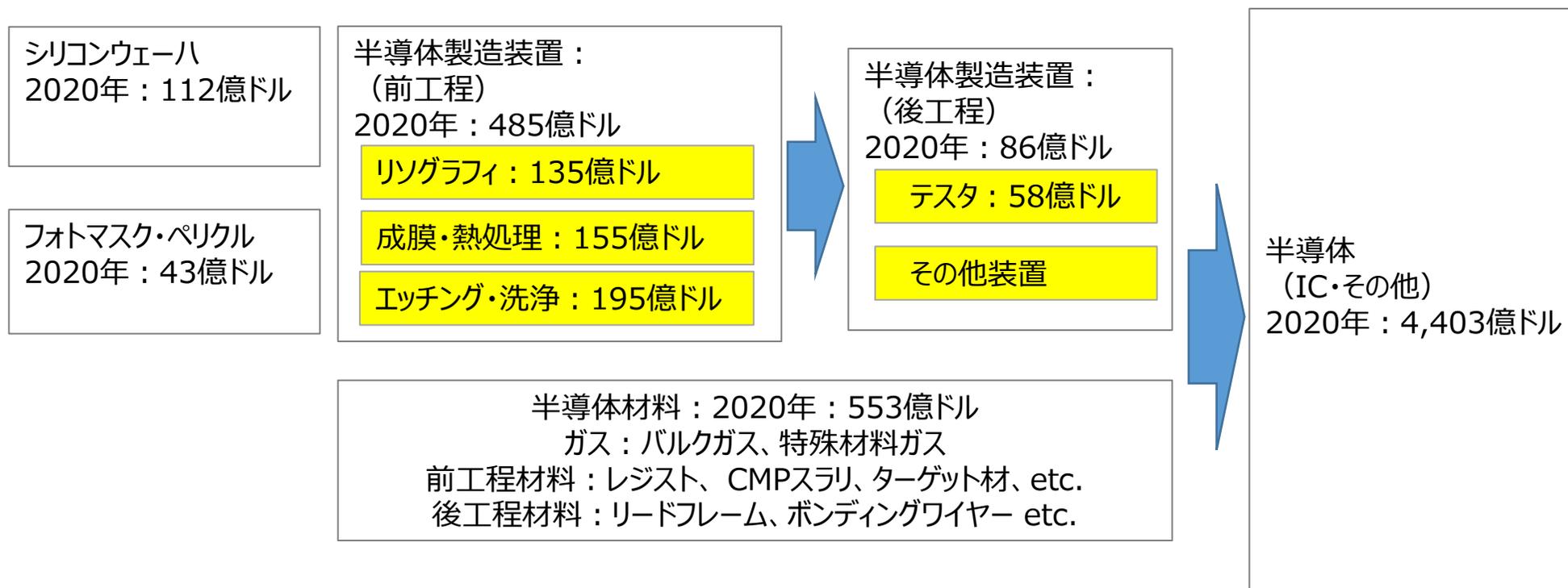
(資料) GLOBAL ELECTRONICS MANUFACTURING SERVICES(EMS) MARKET SIZE, STATUS AND FORECAST 2021-2027(QYRESEARCH社)およびLuxshare HPをもとにNTTアドバンステクノロジーが作成

1-2-7. 半導体製造装置市場

1-2-7-1. 半導体産業におけるサプライチェーン

- 半導体産業におけるサプライチェーンでは、半導体製造装置の市場規模が大きく、その中でもリソグラフィ、成膜・熱処理、エッチング・洗浄という半導体製造時の中心となるプロセス用の装置の市場規模が大きい。

半導体製造の一連の流れと出荷金額に見る市場規模

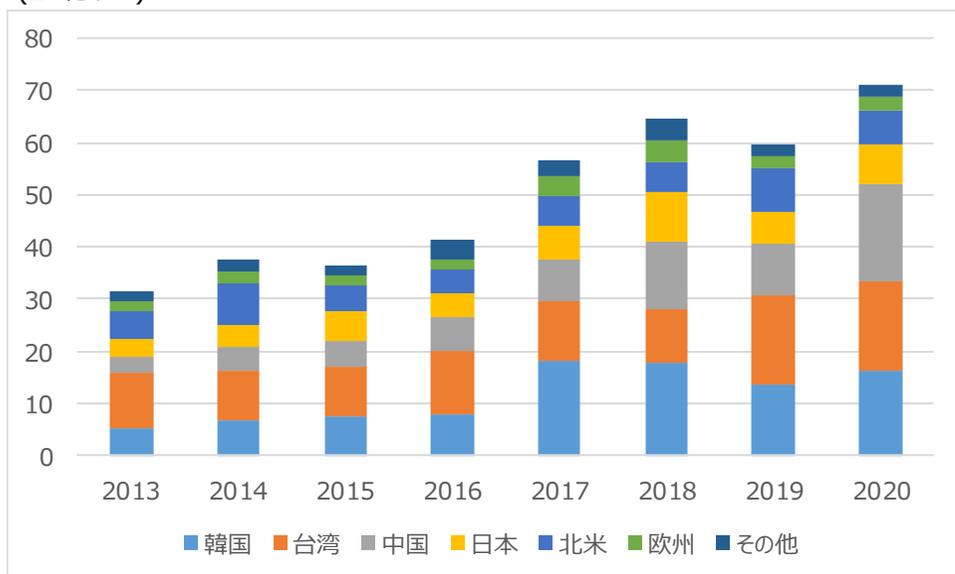


（資料）市場調査レポートの情報をもとにNTTアドバンステクノロジーが作成

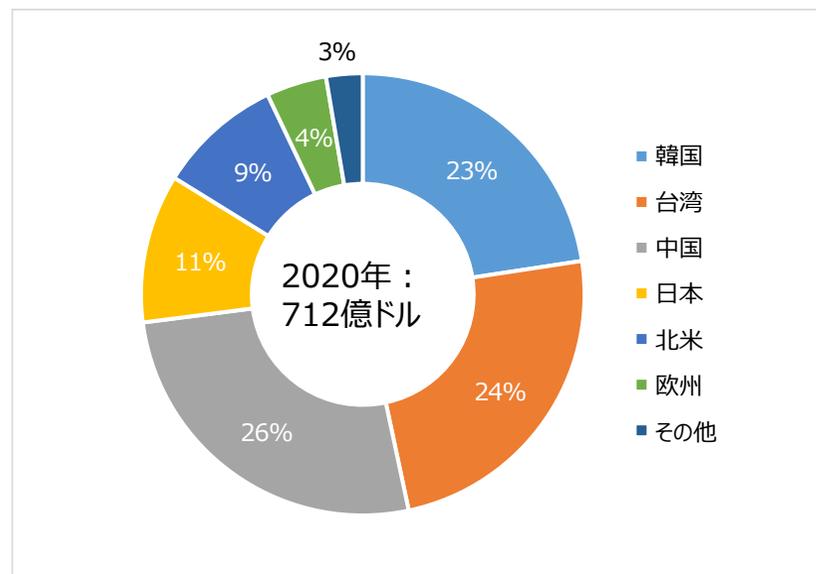
1-2-7-2. 半導体製造装置の世界市場

- 世界総販売額が2019年から19%増え、712億ドルとなった。
- 地域別で初めて**中国が半導体製造装置の最大市場**となった。
- 半導体のスーパーサイクルの影響で製造装置市場も2018年まで伸び、2019年は反動で減少しているが、COVID-19での巣ごもり需要、DX需要で半導体不足が顕在化し、2020年は半導体メーカーの設備投資が増えたためスーパーサイクル前よりも市場が拡大した。

(10億ドル) 半導体製造装置販売金額



半導体製造装置消費地別シェア@2020

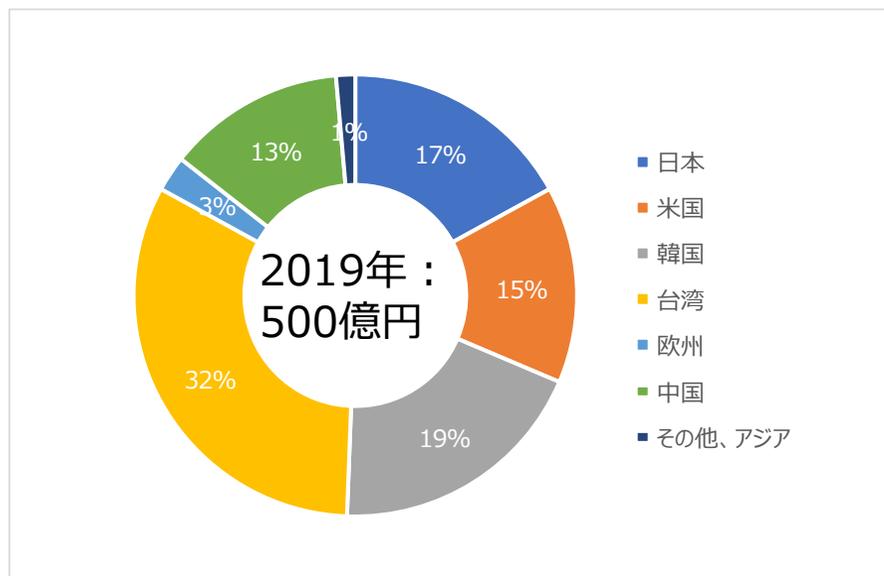


(資料) World Fab Forecast(SEMI)の情報をもとにNTTアドバンステクノロジーが作成

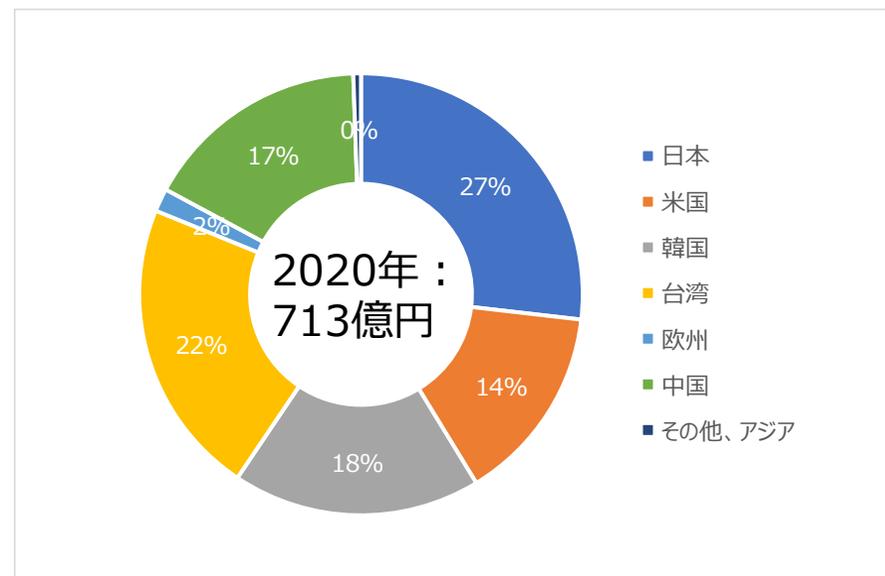
1-2-7-3. 半導体製造 前工程：リソグラフィプロセス マスク/レチクル欠陥装置検査装置市場規模

- 2019年から2020年にかけて42%市場が成長した。
- 消費地別では、日本が大きく伸びた。

マスク/レチクル欠陥検査装置シェア@2019



マスク/レチクル欠陥検査装置シェア@2020

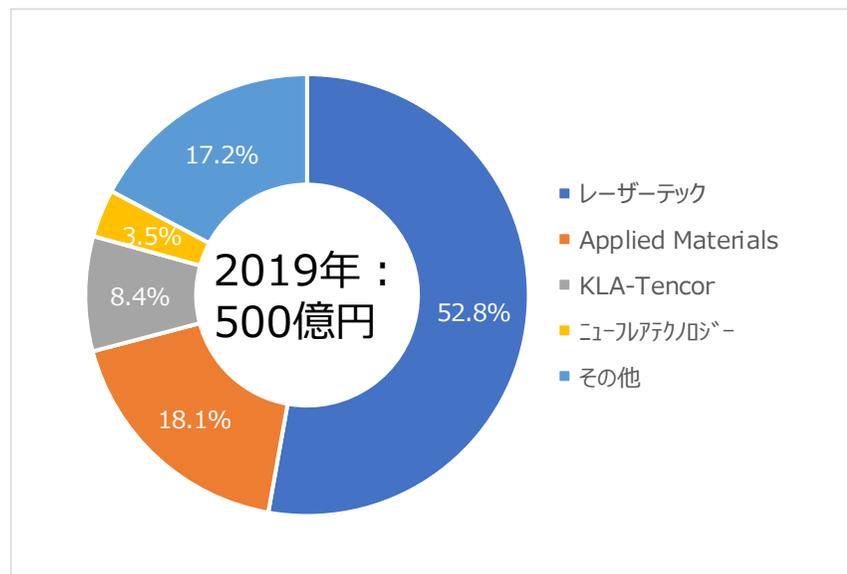


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

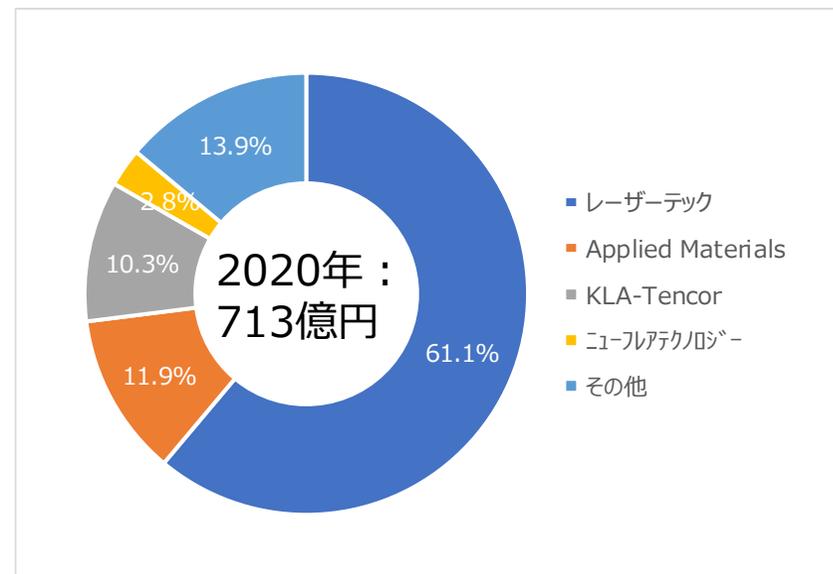
1-2-7-3-1. 半導体製造 前工程：リソグラフィプロセス マスク/レチクル欠陥検査装置市場シェア

- メーカー別ではレーザーテックがシェアを拡大して61.1%となった。レーザーテックがいち早くEUV光源を欠陥検査装置に搭載したことでシェアを拡大していると思われる。

マスク/レチクル欠陥検査装置シェア@2019



マスク/レチクル欠陥検査装置シェア@2020

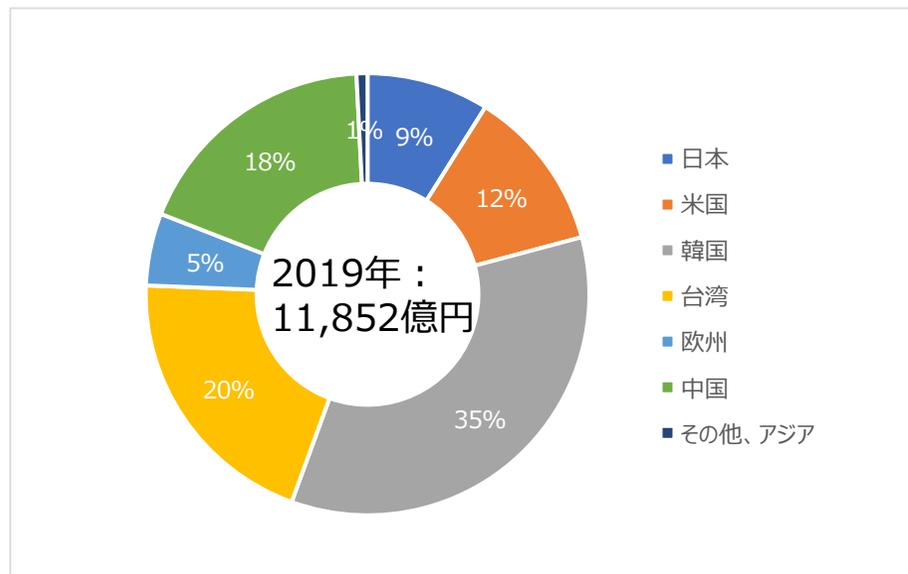


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

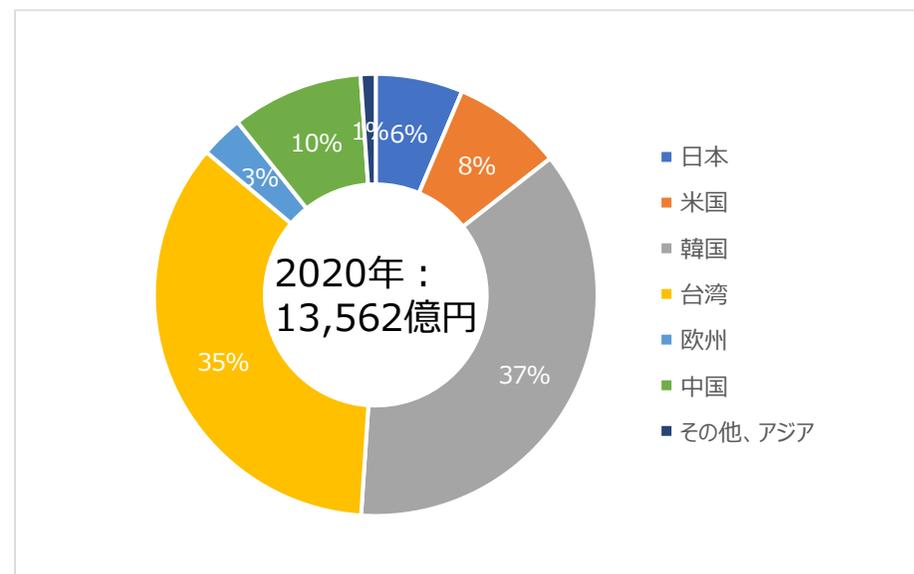
1-2-7-4. 半導体製造 前工程：リソグラフィプロセス ステッパ/スキャン装置市場規模

- 市場規模が14%成長した。
- 消費地別では、台湾が大きく伸び、中国が減少した。

ステッパ/スキャン装置市場シェア@2019



ステッパ/スキャン装置市場シェア@2020

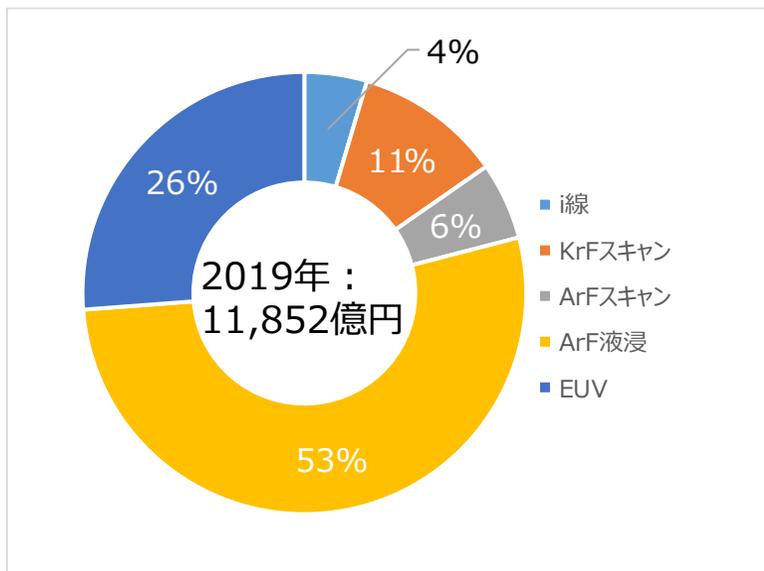


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

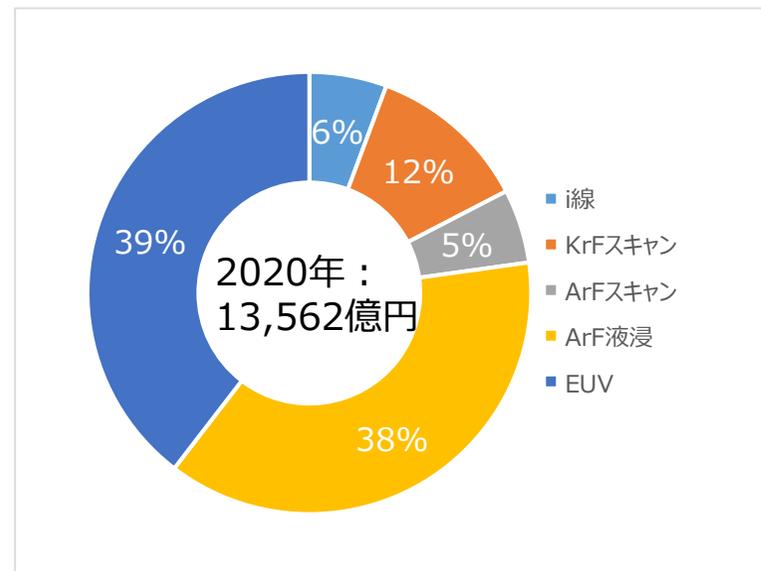
1-2-7-4-1. 半導体製造 前工程：リソグラフィプロセス ステッパ/スキャン装置市場シェア(Technology)

- EUVが26%から39%と大きくシェアを伸ばした。またレガシープロセスノード世代の半導体需要増に対応するためi線、KrFもシェアを拡大した。
- 反対にArF液浸は大きくシェアを落とし、ArFもシェアを落とした。

ステッパ/スキャン装置市場シェア(Technology)@2019



ステッパ/スキャン装置市場シェア(Technology)@2020

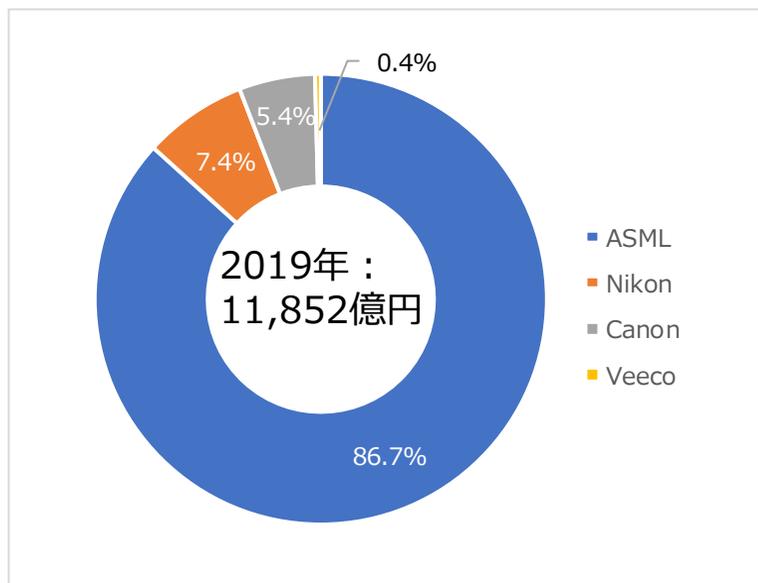


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

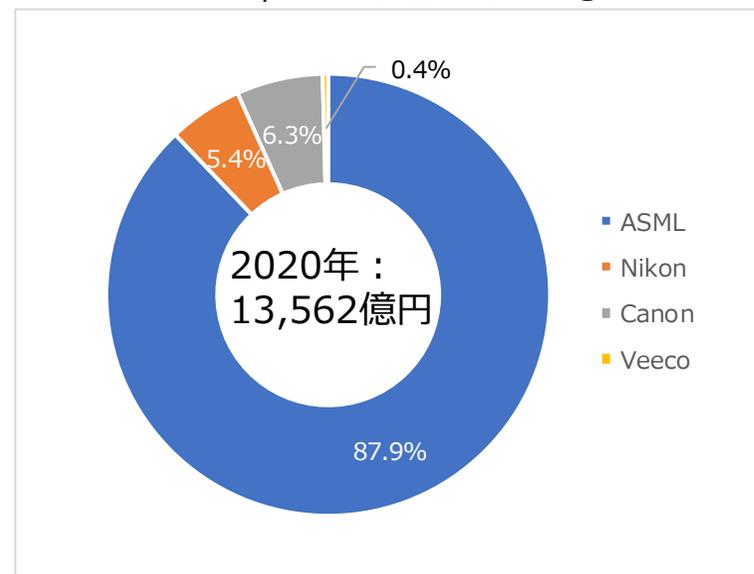
1-2-7-4-2. 半導体製造 前工程：リソグラフィプロセス ステッパ/スキャン装置市場シェア

- TechnologyシェアでEUVが伸びた影響で、EUV露光装置を独占しているASMLがさらにシェアを伸ばした。
- Technologyシェアを伸ばしたi線、KrFに強いCanonがNikonをランキングで逆転した。

ステッパ/スキャン装置市場シェア@2019



ステッパ/スキャン装置市場シェア@2020

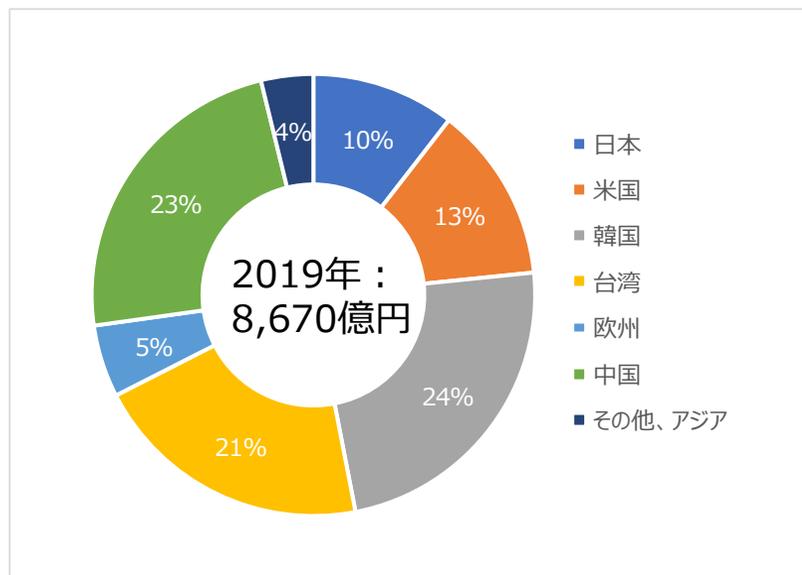


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

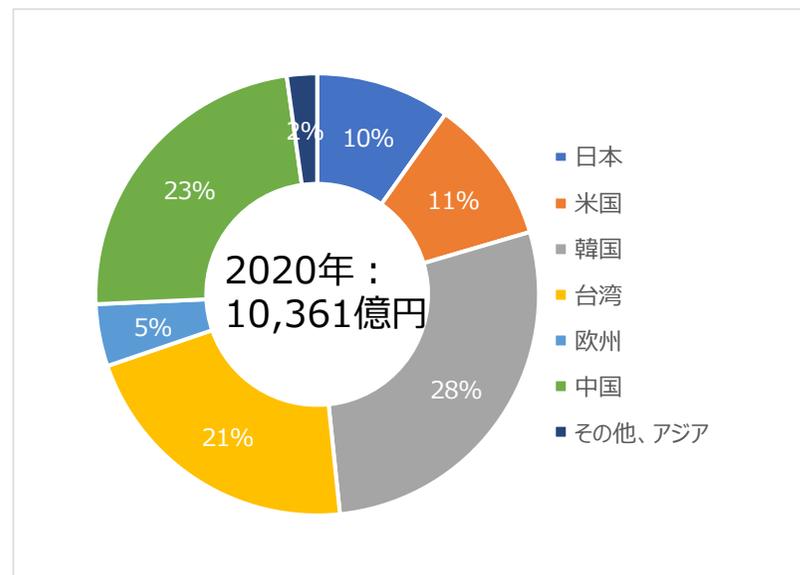
1-2-7-5. 半導体製造 前工程：成膜プロセス CVD/ALD装置市場規模

- 前年から市場規模は成長している。
- 消費地別の割合に大きな変動はなかった。

CVD/ALD装置市場シェア@2019



CVD/ALD装置市場シェア@2020

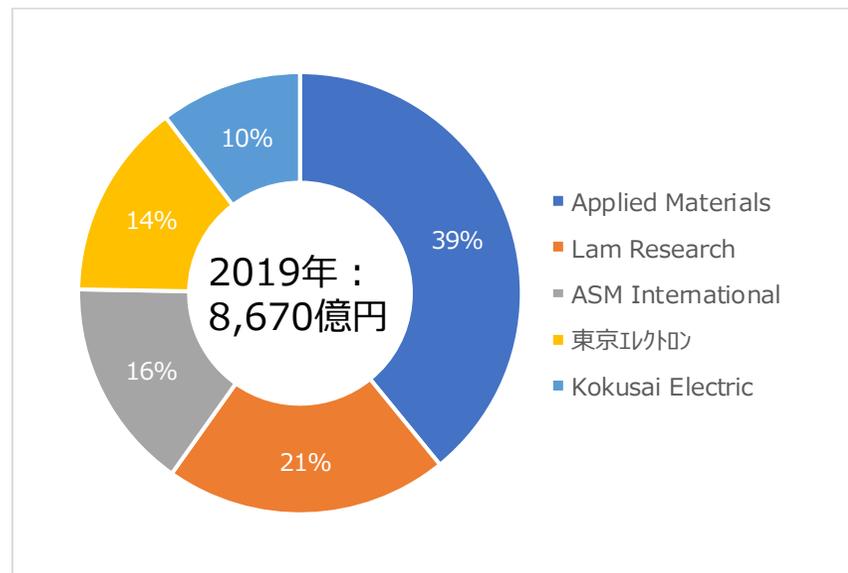


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

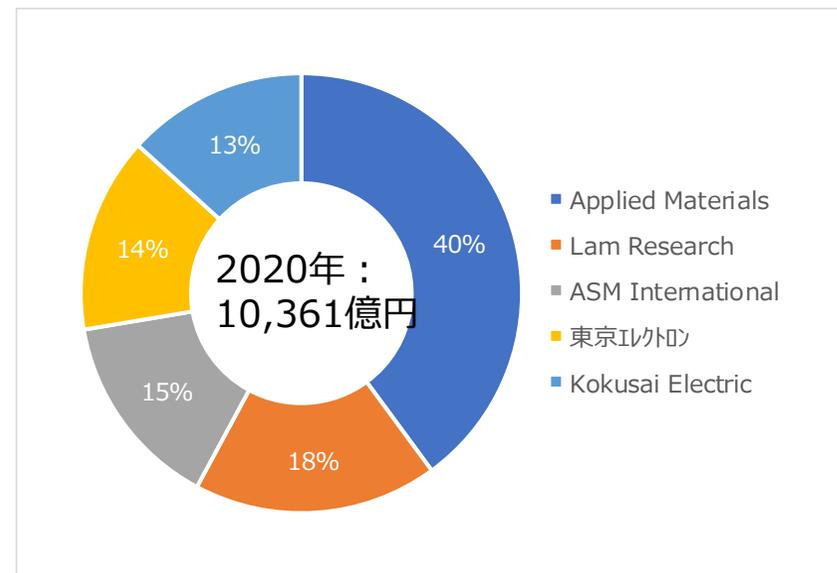
1-2-7-5-1. 半導体製造 前工程：成膜プロセス CVD/ALD装置市場シェア

- シェアトップはApplied Materialsの40%で、メーカー別シェアの順位に変動はなかった。

CVD/ALD装置市場シェア@2019



CVD/ALD装置市場シェア@2020

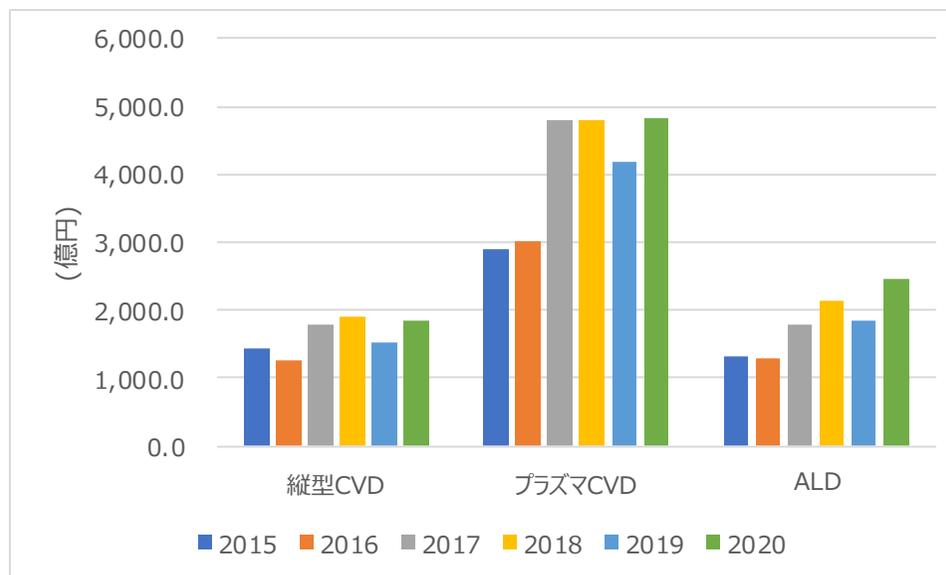


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-5-2. 半導体製造 前工程：成膜プロセス CVD/ALD装置市場規模推移

- 成膜装置の種類別の動向をみると、縦型CVD、プラズマCVDは2019年の落ち込み後、2020年は2018年と同程度まで2020年に回復しているものの横ばいとなっている。
- ALDも2019年に落ち込んでいるが、2018年よりも2020年の市場規模が大きくなっている。その理由は、ALDの精密な成膜、高アスペクト比のトレンチのような難しい形状の成膜などの需要が伸びているためと思われる。

CVD/ALD装置市場規模推移



(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-5-3. 半導体製造 前工程：成膜プロセス 成膜装置の生産拠点の状況

- 日本企業は国内で生産し、販売地での設置、その後のサポートは現地が主流となっている。
- 海外企業は自国内に加えて、需要地に近い地域で生産、販売を行っている。

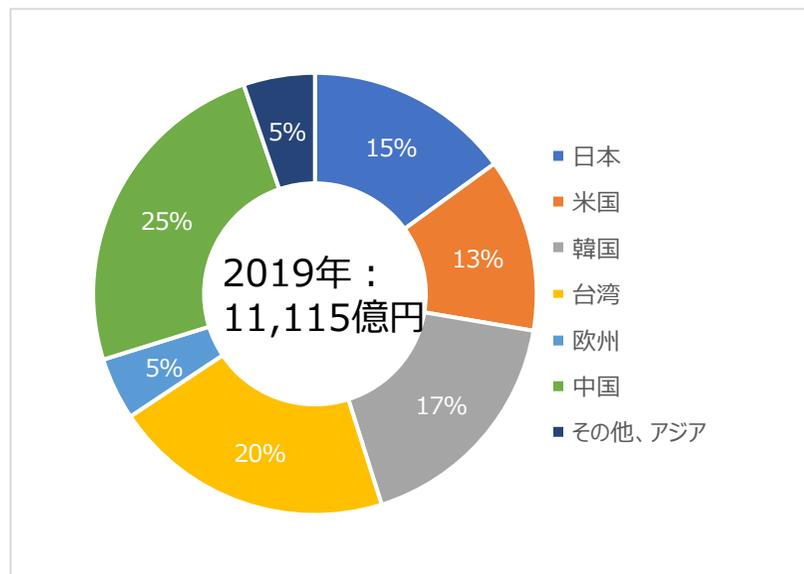
企業国籍	主要企業	生産地
米国	Applied Materials	米国・欧州・シンガポール
米国	LAM Research	米国・欧州・韓国
日本	東京エレクトロン	日本
オランダ	ASM International	欧州・シンガポール

(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

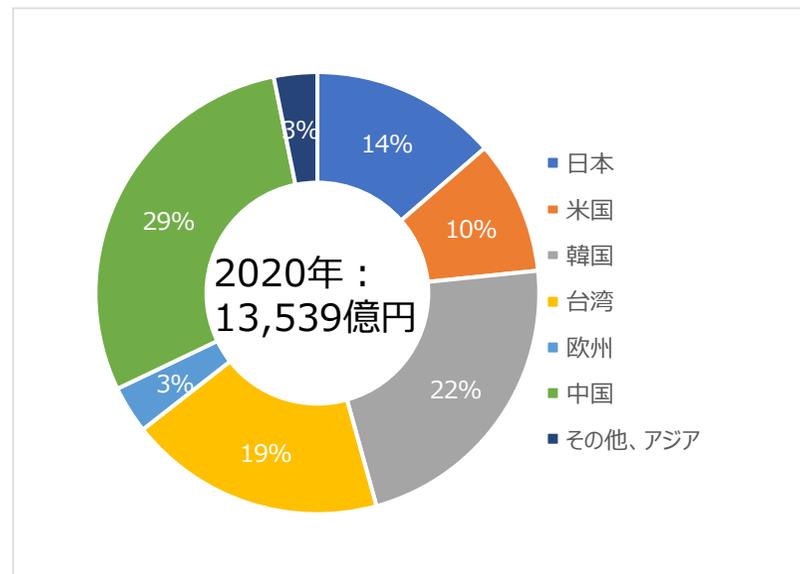
1-2-7-6. 半導体製造 前工程：エッチングプロセス エッチング装置市場規模

- 前年から市場規模は拡大した。
- 消費地別では中国がトップの29%となった。
- 韓国が3位から2位に、台湾が2位から3位にランキングが変動した。

エッチング装置市場シェア@2019



エッチング装置市場シェア@2020

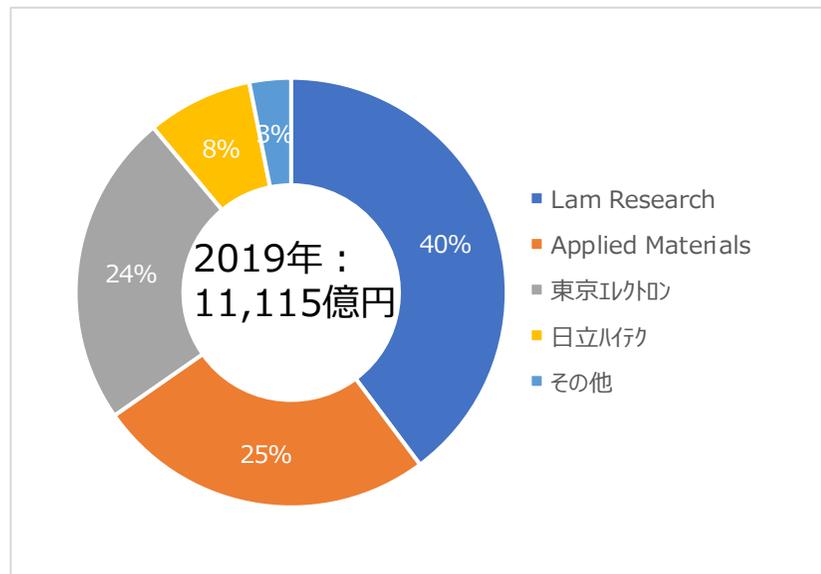


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

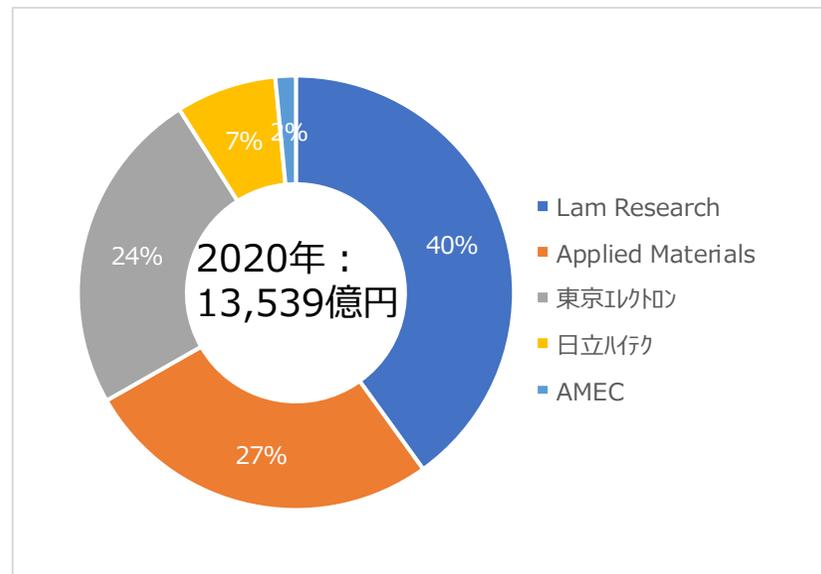
1-2-7-6-1. 半導体製造 前工程：エッチングプロセス エッチング装置市場シェア

- メーカー別シェアに大きな変動はなかった。
- トップのLam Researchが40%、次いでApplied Materialsが27%であった。

エッチング装置市場シェア@2019



エッチング装置市場シェア@2020

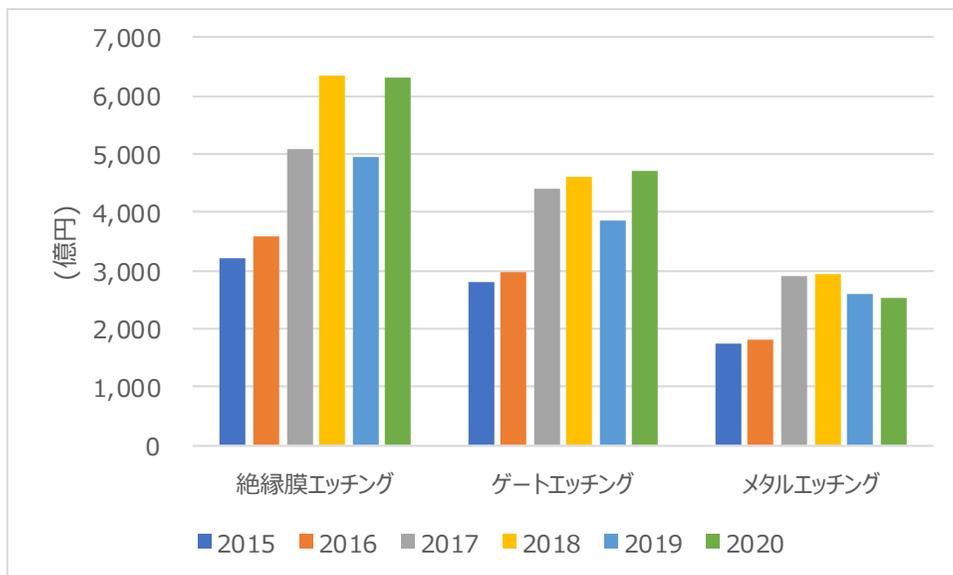


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-6-2. 半導体製造 前工程：エッチングプロセス エッチング装置種別毎の市場規模推移

- NANDメモリ製造に使われる絶縁膜エッチング装置が最も市場規模が大きい。
- 2019年に落ち込んだ市場が、絶縁膜エッチング装置とゲートエッチング装置は2020年には2018年と同程度まで回復した。しかし、メタルエッチング装置は2019年よりも減少した。

エッチング装置 装置種別毎の推移



【絶縁膜エッチング装置】：NANDフラッシュメモリ向けメモリホールの形成のため、高アスペクト比性能が求められる。LAM Researchではエッチングチャンバ内でアッシングを実施することも可能で、プロセス時間を大幅に短縮する。

【ゲートエッチング装置】：微細化が進むことで3次元構造を原子スケールで設計通りに加工することが求められる。さらに、プラズマ加工におけるデバイス材料表面へのダメージ形成(材料物性変化)を最小化することも同時に求められる。近年ではALDを応用した原子層エッチングも盛んに研究されている。

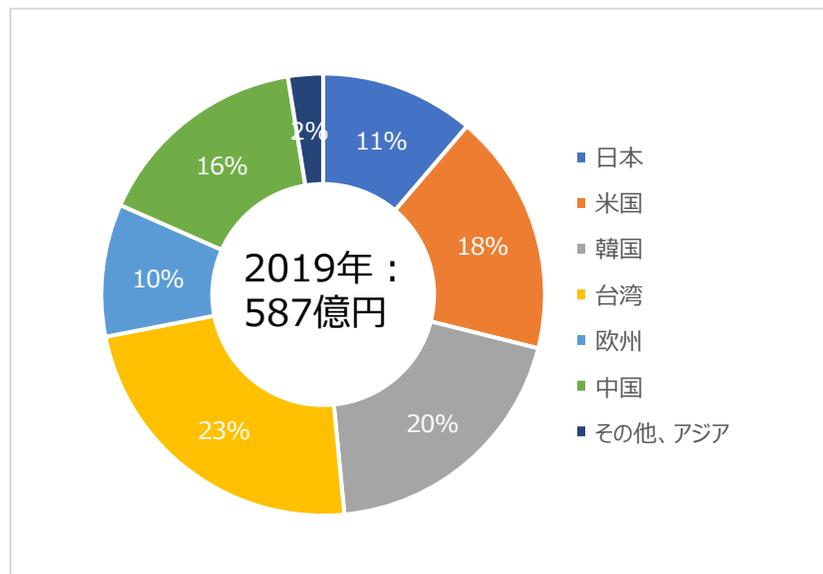
【メタルエッチング装置】：導線や電気配線を形成するエッチング装置。低圧化、Pt、Ir等の強誘電体材料への対応、High-kメタルゲートへの対応、ドライエッチングが難しいとされるCu配線への対応が求められる。近年ではメタルハードマスクのパターン形成用途でも利用される。

(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-6. 半導体製造 前工程：エピタキシャル エピタキシャル成長装置市場規模

- 前年から市場規模は大きくなっており、699億円になった。
- 消費地別シェアに大きな変動はなく、台湾がトップで23%であった。

エピタキシャル成長装置市場シェア@2019



エピタキシャル成長装置市場シェア@2020

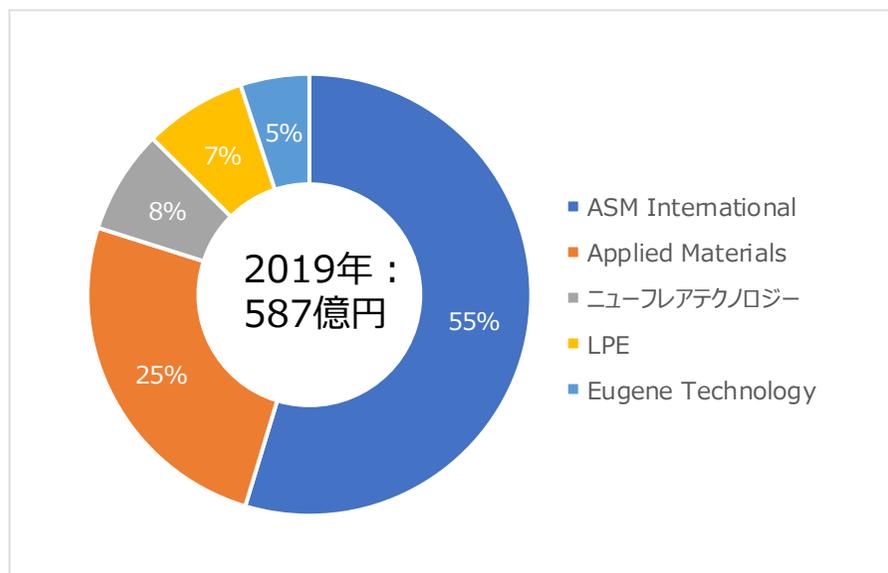


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

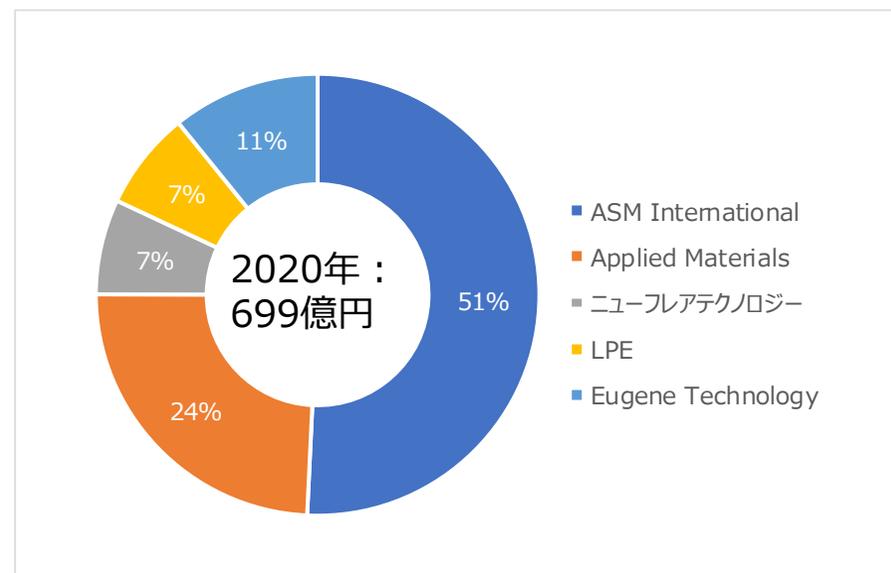
1-2-7-6-1. 半導体製造 前工程：エピタキシャル エピタキシャル成長装置市場シェア

- 前年からシェア1位～3位のシェアが減少し、シェア5位のEugene Technologyがシェアを倍増させた。2020年にEugene Technologyが欧州でのシェアを急激に伸ばした。

エピタキシャル成長装置市場シェア@2019



エピタキシャル成長装置市場シェア@2020



(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-7. 半導体製造 後工程： ダイシングソー市場規模

- 前年から市場が大きく拡大した。
- 消費地別シェアトップの中国がシェアを増やして26%になった。また、シェア5位だった韓国が16%になり、日本、台湾にシェアがならんだ。

ダイシングソー市場シェア@2019



ダイシングソー市場シェア@2020

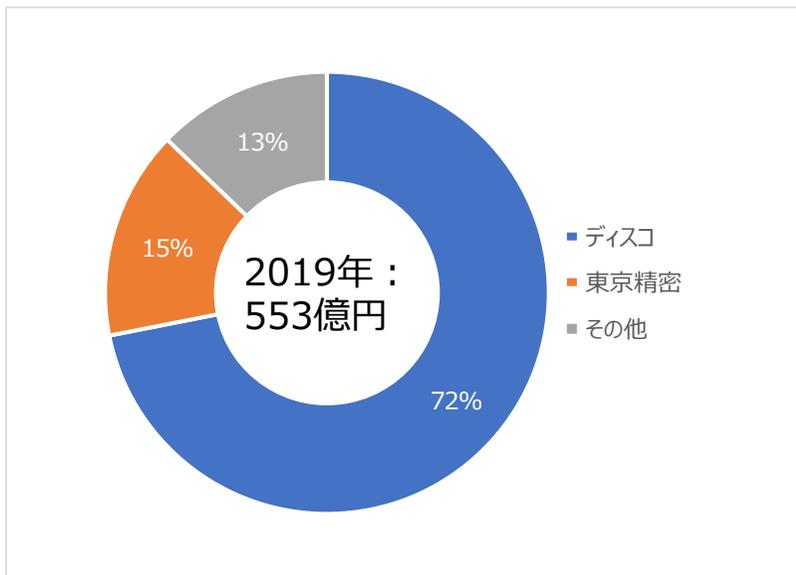


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

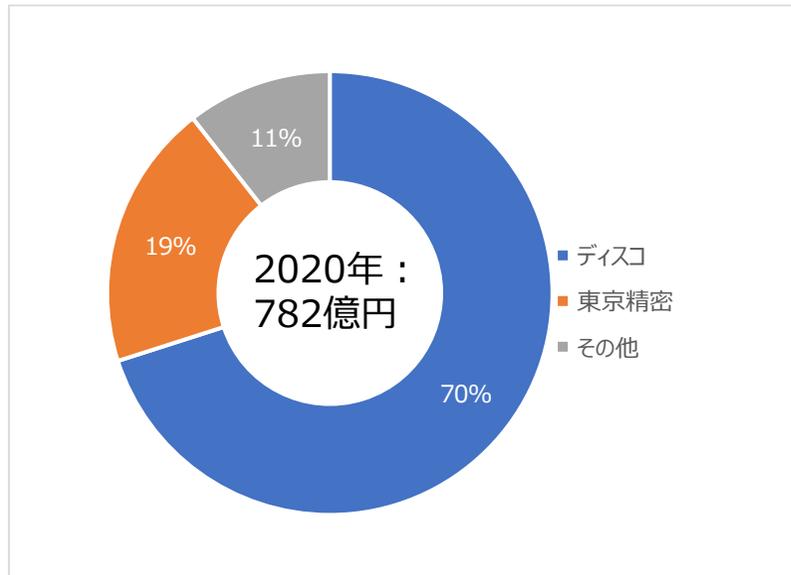
1-2-7-7-1. 半導体製造 後工程： ダイシングソー市場シェア

- メーカー別シェアではディスコが強く、70%のシェアを持っている。
- 東京精密もシェアを伸ばし、19%となり、上位2社で89%を占めている。

ダイシングソー市場シェア@2019



ダイシングソー市場シェア@2020

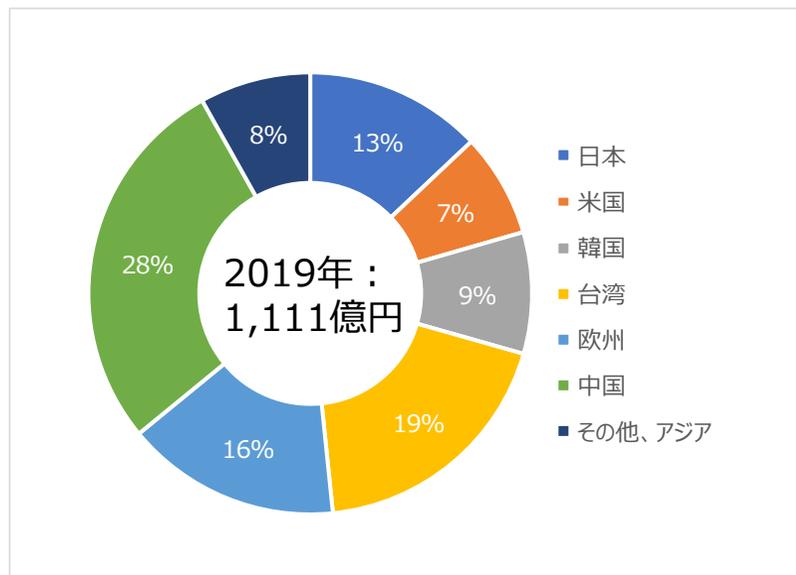


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

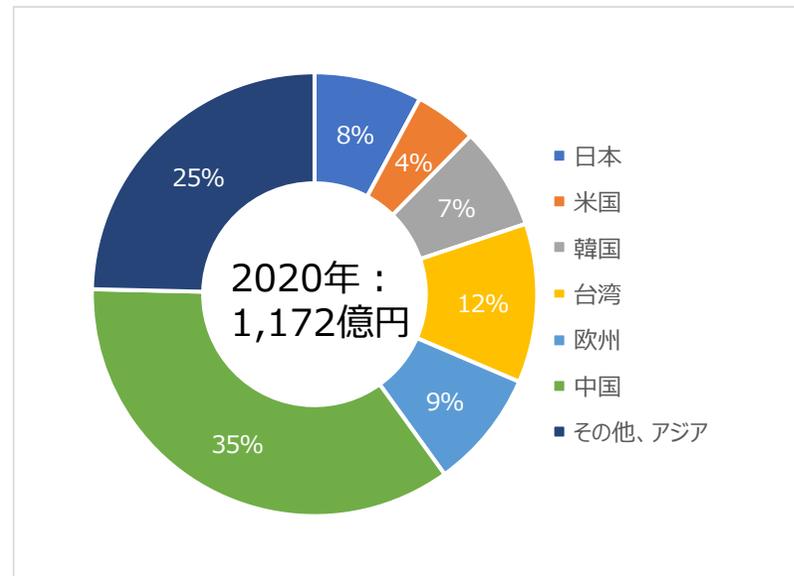
1-2-7-8. 半導体製造 後工程： ダイ/チップボンディング装置市場規模

- 前年から市場規模は増えているが、小幅な拡大となっている。
- 消費地別シェアでは、中国以外は軒並みシェアが減少し、その他アジアが大幅に増加した。
- トップの中国はシェアを伸ばし、35%になった。

ダイ/チップボンディング装置市場シェア@2019



ダイ/チップボンディング装置市場シェア@2020

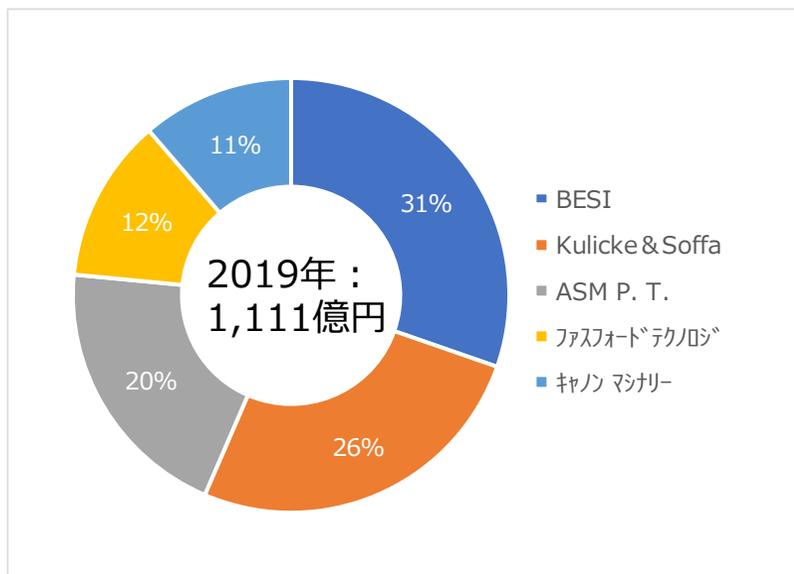


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-8-1. 半導体製造 後工程： ダイ/チップボンディング装置市場シェア

- メーカー別シェアトップのBESIがシェアを伸ばし、40%となった。
- シェアの変動が激しく、昨年と比べてChangChuanとヤマハ発動機(新川)が伸びた。

ダイ/チップボンディング装置市場シェア@2019



ダイ/チップボンディング装置市場シェア@2020

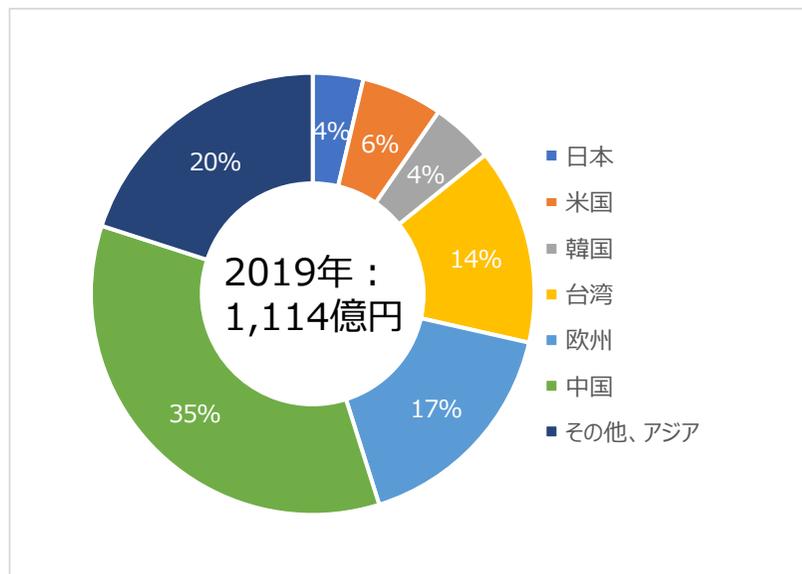


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

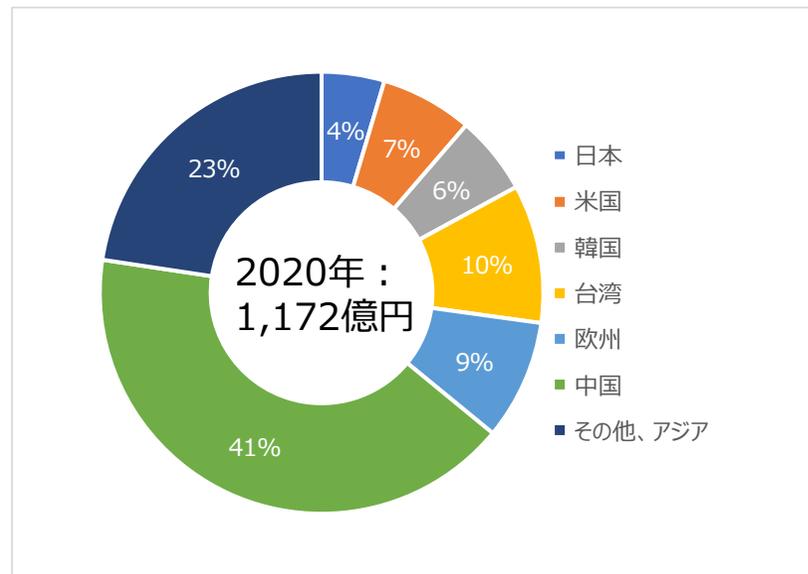
1-2-7-9. 半導体製造 後工程： ワイヤボンディング装置市場規模

- 前年から市場規模は増えているが、小幅な拡大となっている。
- 消費地別シェアでは、中国、米国以外は軒並みシェアが減少した。
- トップの中国はシェアを伸ばし、41%になった。

ワイヤボンディング装置市場シェア@2019



ワイヤボンディング装置市場シェア@2020

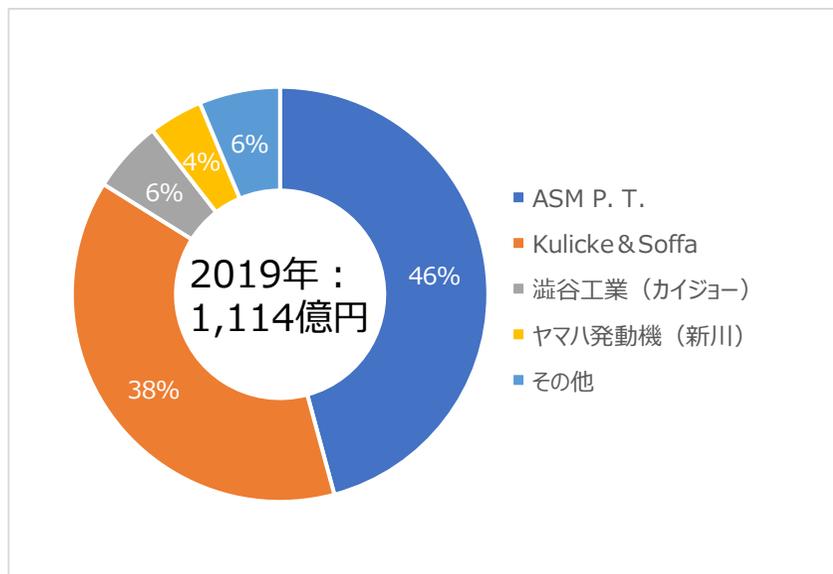


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-9-1. 半導体製造 後工程： ワイヤボンディング装置市場シェア

- ASM P.T.とKulicke & Soffaが2強であり、この2社で85%を占めている。
- 澁谷工業(カイジョー)がシェアを大きく減らした。

ワイヤボンディング装置市場シェア@2019



ワイヤボンディング装置市場シェア@2020



(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

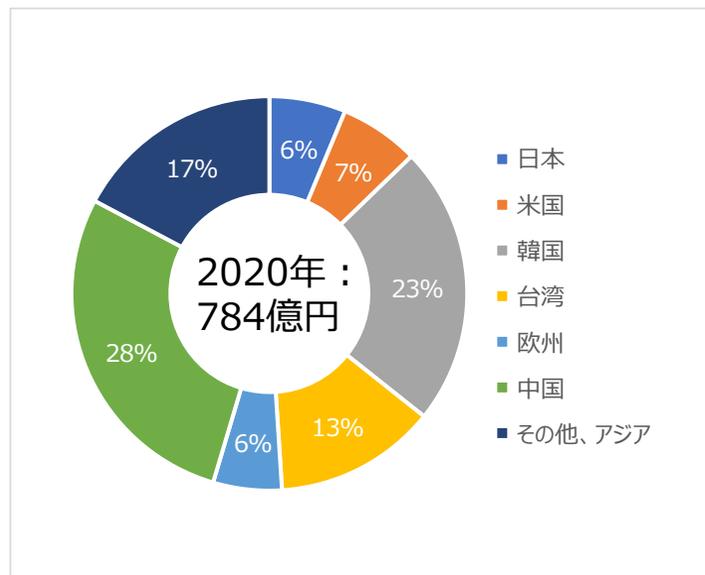
1-2-7-10. 半導体製造 半導体テスト： ICハンドラ市場規模

- 市場規模はわずかだが増加し784億円になった。
- 消費地別シェアでは、中国と韓国が増加し、それぞれ28%、23%になり、1位、2位。
- 台湾はシェアを減らし、2位から4位になった。

ICハンドラ市場シェア@2019



ICハンドラ市場シェア@2020

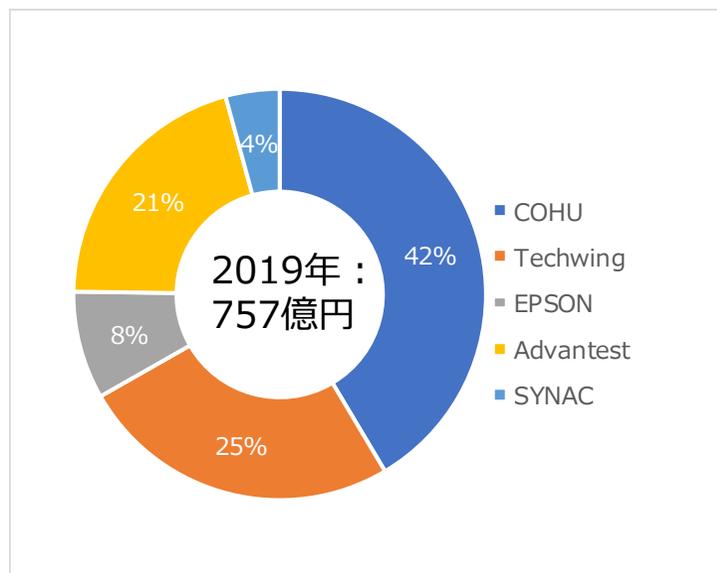


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

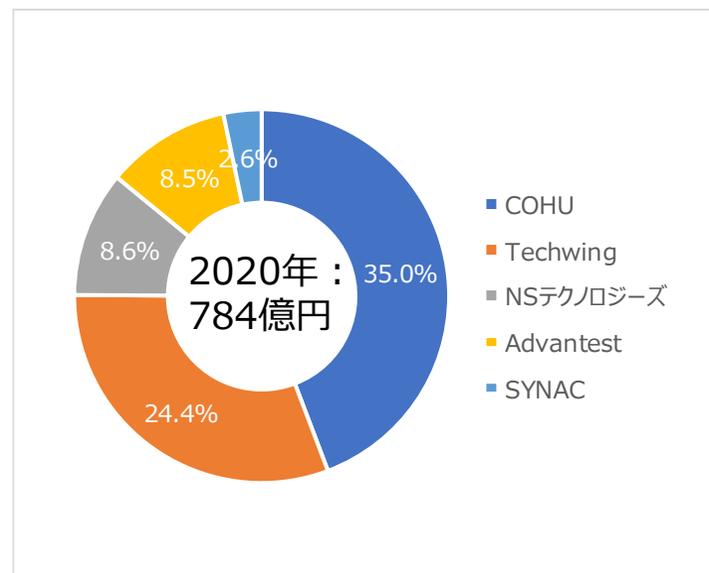
1-2-7-10-1. 半導体製造 半導体テスト： ICハンドラ市場シェア

- メーカー別シェアでは、COHUがトップで35%、次いでTechwingの24.4%、NSテクノロジーズ(旧EPSON)の8.6%となった。
- Advantestがシェアを下げた。

ICハンドラ市場シェア@2019



ICハンドラ市場シェア@2020

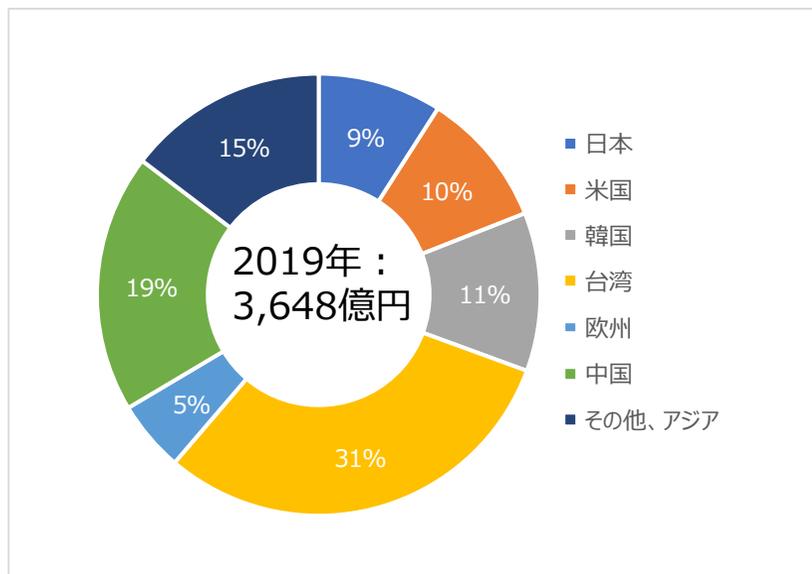


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

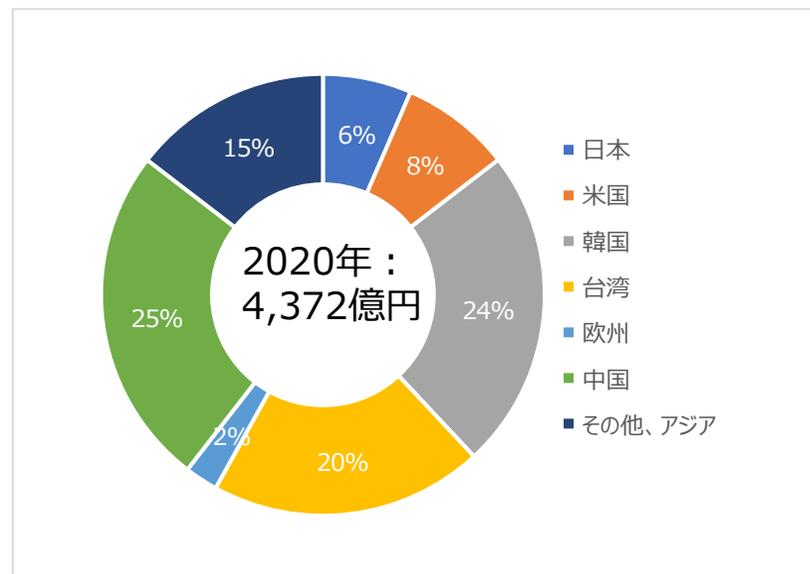
1-2-7-11. 半導体製造 半導体テスト： ICテスト市場規模

- 前年から市場規模は拡大し、4,372億円となった。
- 消費地別シェアでは、韓国が大幅に増加し、1位の中国25%に次ぐ24%となった。
- 台湾はシェアを下げた。

ICテスト市場シェア@2019



ICテスト市場シェア@2020

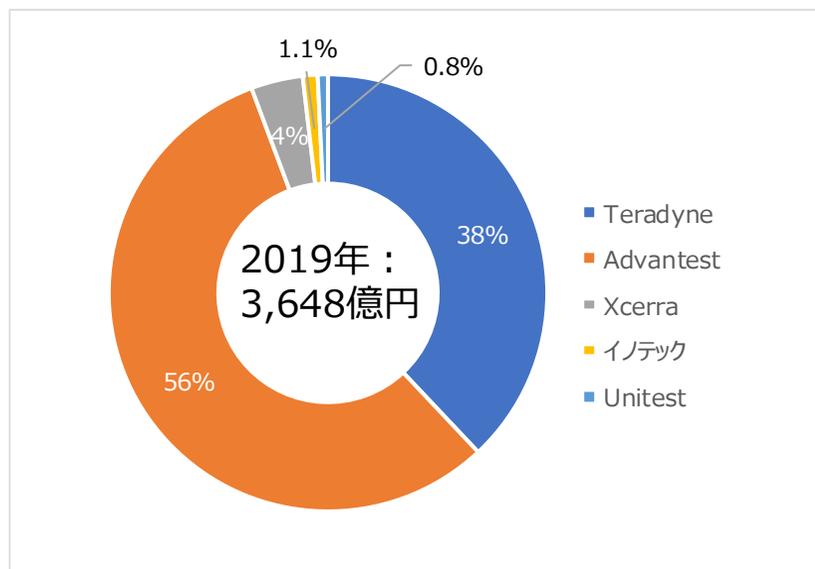


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

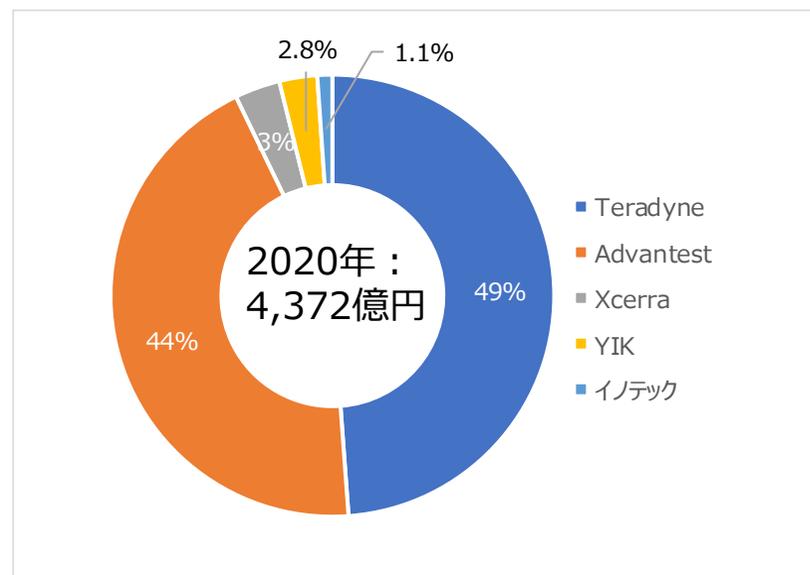
1-2-7-11-1. 半導体製造 半導体テスト： ICテスト市場シェア

- ICテスト市場は、TeradyneとAdvantestの2強で93%を占める寡占市場となっている。
- 2020年はTeradyneがシェア伸ばし、49%で1位となった。

ICテスト市場シェア@2019



ICテスト市場シェア@2020



(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-7-11-2. 半導体製造 半導体テスト： ICテスト生産拠点の状況

- Advantestは日本国内の2工場の他に、韓国、マレーシアの需要地に近い場所で生産を行っている。
- Teradyneは熊本工場でも一部は組立は行っているが、基板の生産は中国・マレーシアのEMSに製造委託している。
- XcerraはアジアでEMSを活用している。

企業国籍	主要企業	生産地
米国	Teradyne	中国・マレーシアのEMSに製造委託
	Xcerra	アジアで製造委託
日本	Advantest	群馬工場、仙台工場、Cheonan(韓国)、Penang(マレーシア)

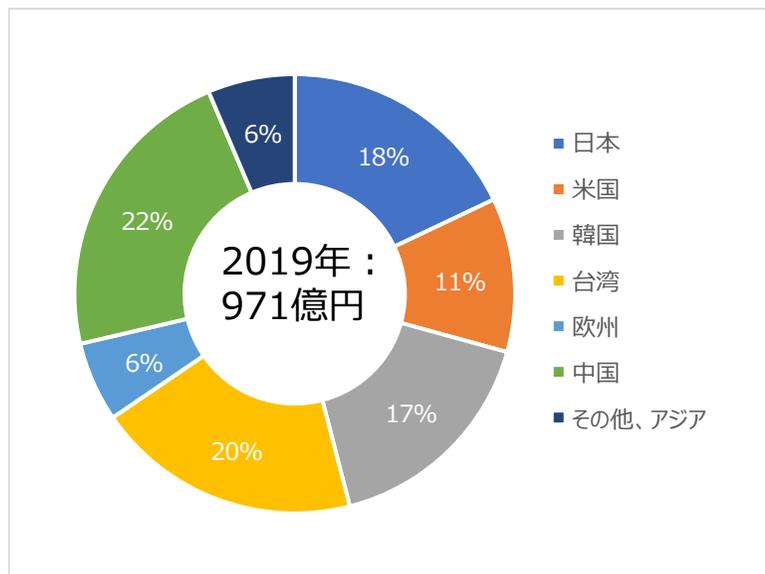
赤字は新設、もしくは新規製造委託

(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

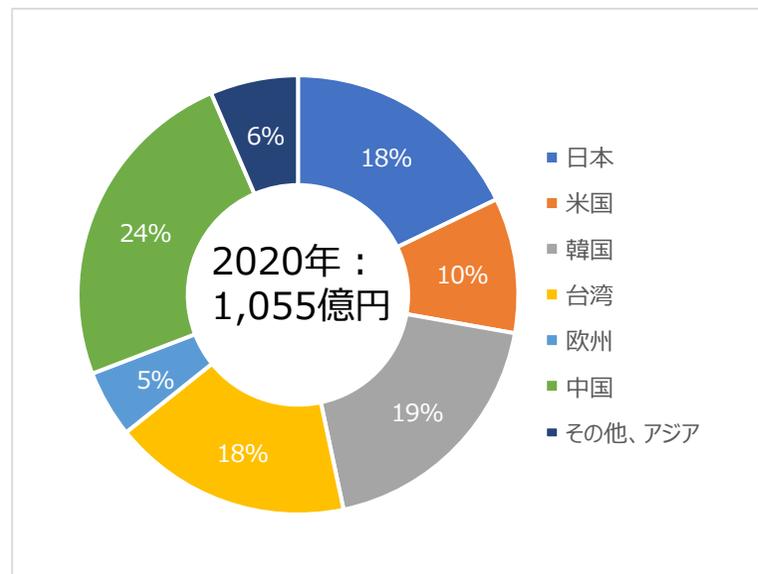
1-2-7-12. 半導体製造 半導体テスト： ウェーハプローバ市場規模

- 前年から市場規模は拡大し、1,055億円になった。
- 消費地別シェアに大きな変動はなかった。
- シェアトップは中国の24%、次いで韓国の19%、日本・台湾の18%であった。

ウェーハプローバ市場シェア@2019



ウェーハプローバ市場シェア@2020

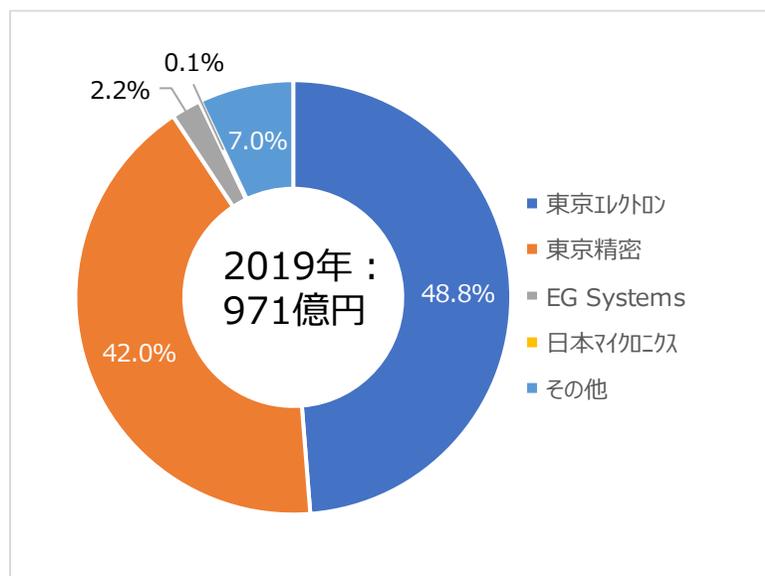


(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

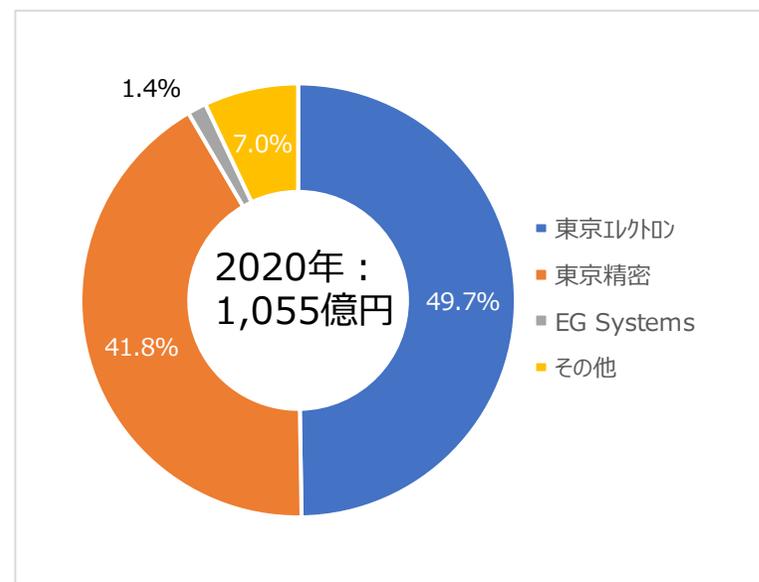
1-2-7-12-1. 半導体製造 半導体テスト： ウェーハプローバ市場シェア

- メーカー別シェアに大きな変化はなかった。
- ウェーハプローバ市場は、東京エレクトロンと東京精密の2強で90%を占めている。

ウェーハプローバ市場シェア@2019



ウェーハプローバ市場シェア@2020



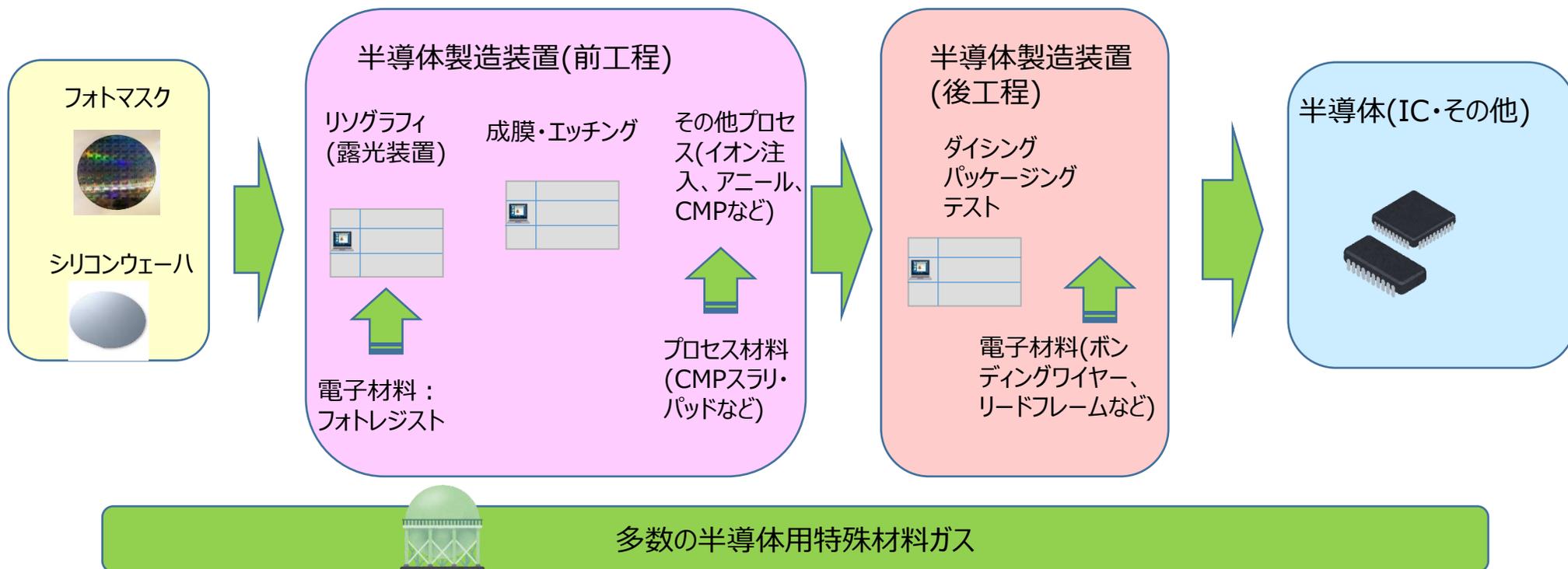
(資料) 世界半導体製造装置・試験/検査装置市場年鑑2021(GNC社)をもとにNTTアドバンステクノロジーが作成

1-2-8. 半導体材料市場

半導体製造のフローチャート

- 半導体材料とは、半導体製品を製造するプロセス、特に半導体ウェーハのパターン化に使用される材料のことである。
- シリコンウェーハ、ウェットケミカル、工業用ガス、化学機械研磨（CMP）スラリーおよびパッド、フォトマスク、スパッターターゲット、半導体用フォトレジスト、EUV用フォトレジスト、静電チャック、ICリードフレームなどの材料が一般的に半導体材料として使用される。

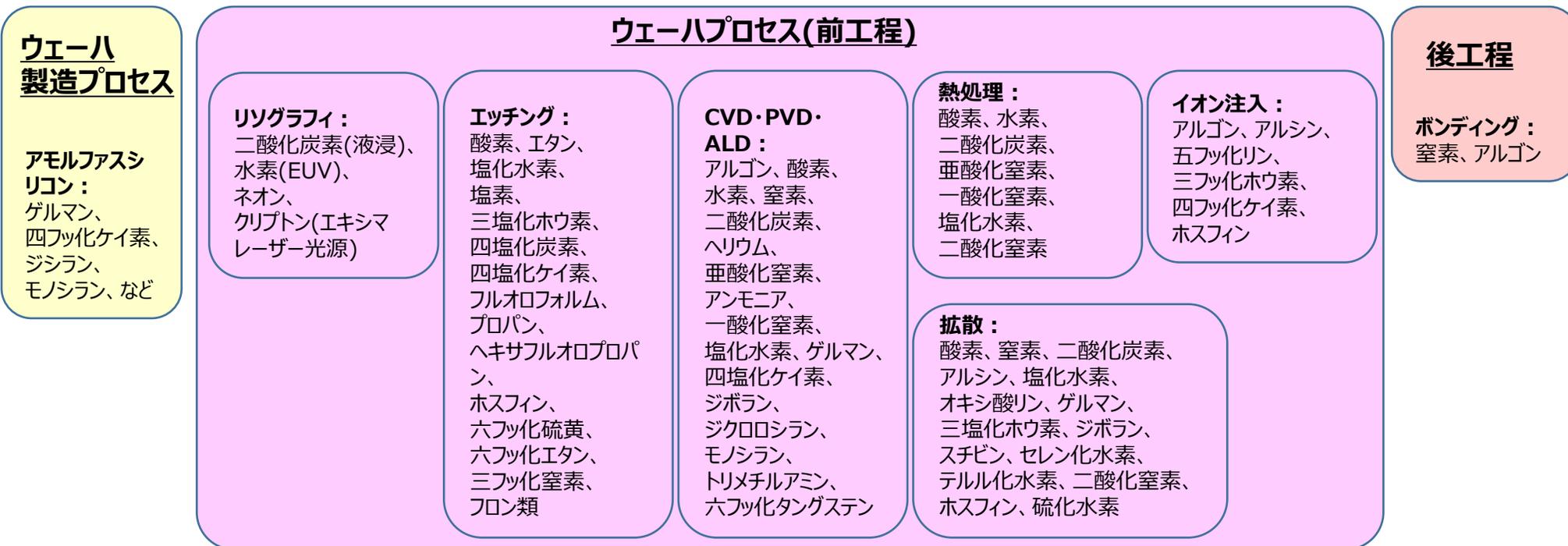
半導体製造のプロセスフロー



半導体製造のフローチャート：主なプロセスと材料（ガス）

- 半導体の製造は複雑であり、プロセス全体でアルゴン、酸素、水素などの30を超える異なるガスが必要。
- 例えば、モノシランは、モノシラン絶縁膜やモノシラン反射防止層の成膜に、三フッ化窒素(NF3)は、プロセスチャンバ壁からのデポジットを洗浄するための洗浄ガスとして、六フッ化タングステン(WF6)は半導体をパターン化して金属接点とゲートを形成する際の化学蒸着プロセスを通じて、ClF3はツールチャンバーの洗浄ガスとして、それぞれ使用される。

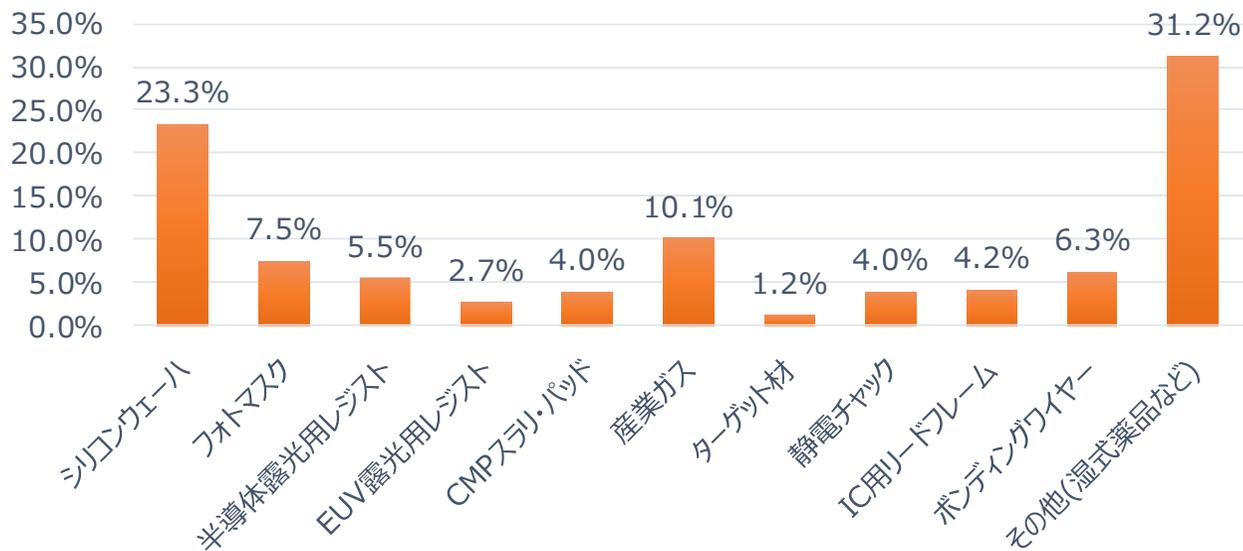
半導体製造のプロセスフローと主な材料ガス



半導体材料市場における変化

- 半導体材料市場は、材料価格の高騰やCOVID-19のパンデミックの突然の発生など、いくつかの不確実性があるものの、ここ数年は緩やかに成長している。
- 特にシリコンウェーハは、2020年の売上シェアは23.3%であり、2021-2026年の予測期間中のCAGRは6.6%となる見込みである。
- 2020年に8インチウェーハの購入価格は前年比で約20%上昇した。これは、世界最大のウェーハメーカーである信越化学工業が2021年4月に20%の値上げを発表した後のことであった。韓国の唯一のウェーハサプライヤーであるSK Siltronは、長期契約のためにすぐに値上げしなかったが、今後の値上げは避けられないと見られている。

2020年材料別売上シェア



主要材料・主要部品に関する補足：原材料（金属）価格推移

- ボンディングワイヤーは、金やアルミニウムなどの材料で作られた細いワイヤーで、半導体とシリコンチップの間に電氣的相互接続を作成するプロセス(ワイヤーボンディング)で使用される。
- 金価格は2020年中旬以降は増加傾向が収まっていた。しかし、ウクライナ情勢の影響で2022年3月以降は高騰し始めている(LMEデータとしては今後反映される)。



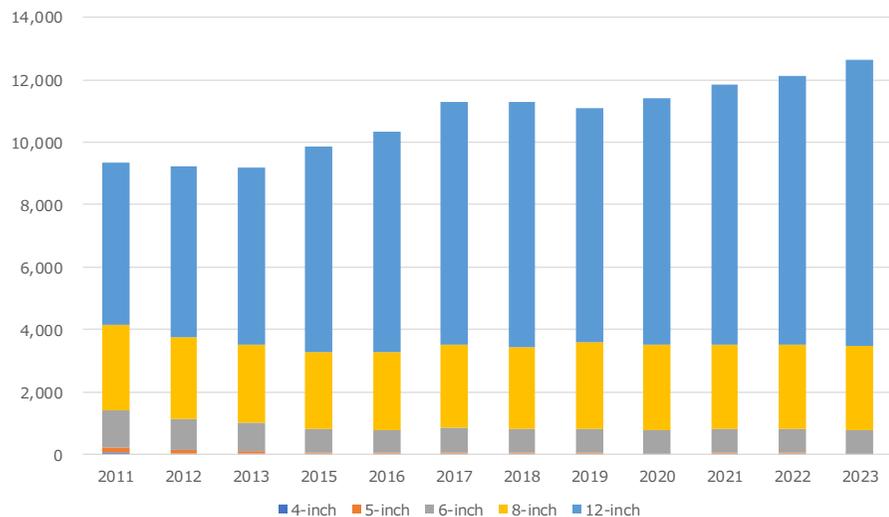
出典：R02報告書、LME(London Metal Exchange)

主要原料・主要部品の販売状況

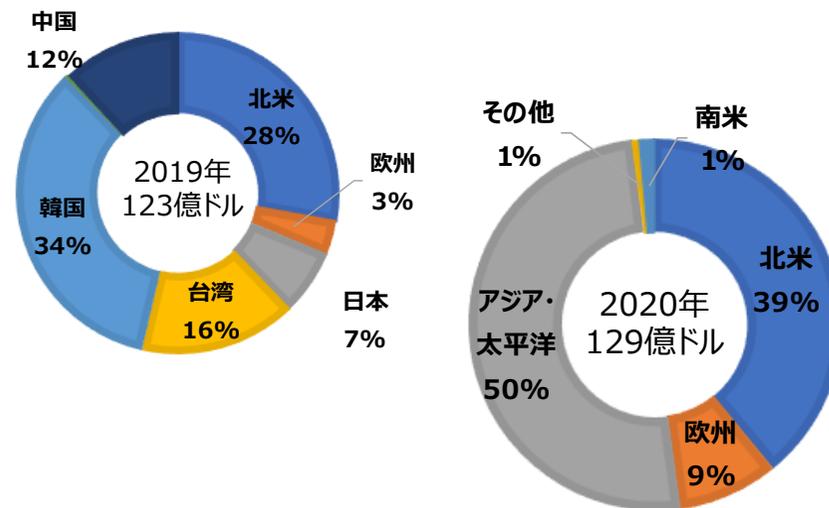
電子材料-シリコンウェーハ

- Covid-19の大流行の後、状況の改善と最終用途市場の成長に伴い、シリコンウェーハの需要は成長した。例えば、SEMIシリコンマニュファクチャラーグループ（SMG）は、2021年の第2四半期に世界のシリコンウェーハ面積の出荷量が6%増加して35億3400万平方インチになり、第1四半期の過去最高を上回ったと報告した。

百万平方インチ シリコンウェーハ消費面積推移



シリコンウェーハ消費地域別シェア



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (面積推移と2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

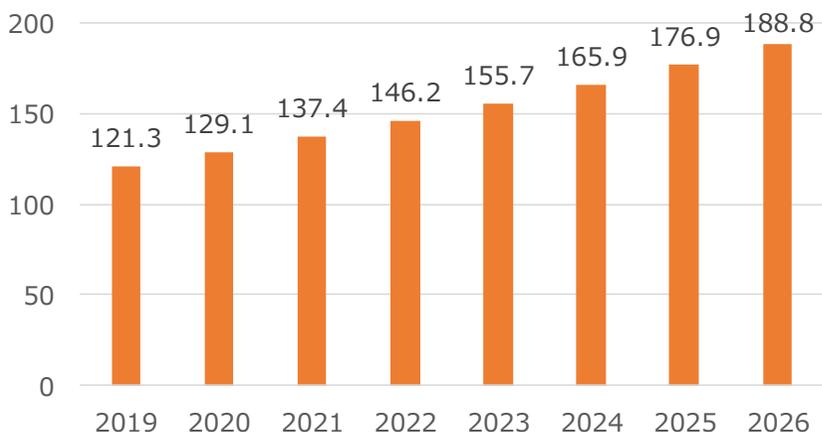
主要原料・主要部品の販売状況

電子材料-シリコンウェーハ

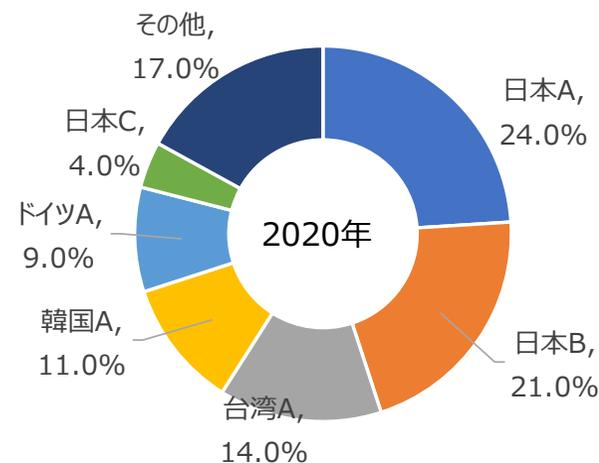
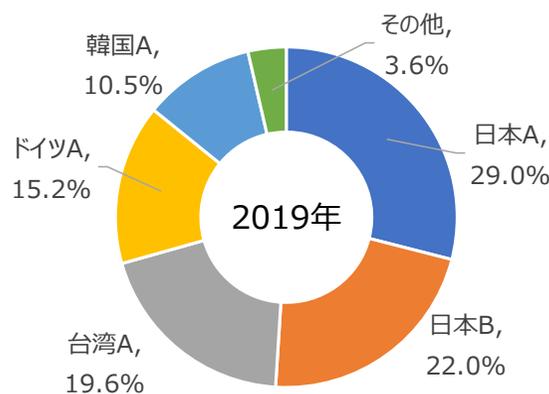
- 2020年のシリコンウェーハ市場規模は2019年から6.4%増で、今後も順調に増加すると予想されている。
- 2019年と2020年で情報源が異なるため、ベンダシェアの数値は多少増減しているが、シェアの傾向は大きく変わっていない。

シリコンウェーハ世界市場規模推移

億ドル



シリコンウェーハ売上高ベンダシェア



(出典) 各種情報源よりNTTアドバンステクノロジ株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-シリコンウェーハ

- SUMCOは、高精度の製造技術と品質管理システムを使用して、超フラットで超クリーンな製品を製造している。2021年に、シリコンウェーハ生産能力増強のために新株を通じて11億米ドルを調達すると発表した。なお、米国とインドネシアでは200mmのみ製造。
- 2020年にGlobal WafersはSiltronicの買収を発表したが、2022年2月時点でドイツ政府からの承認が得られず、買収は実現しなかった。一方、同じ2020年にはSK Siltron が独 Dupont の SiC ウェーハ部門を買収した。

半導体シリコンウェーハにおけるサプライチェーン：各国企業の生産地

国・地域	主要企業	生産地
日本	信越グループ ^{*1}	日本、マレーシア、台湾、英国、米国
	SUMCO	日本、台湾、米国、インドネシア
	フエロテック	中国
	昭和電工	日本
韓国	SK Siltron	韓国
台湾	Wafer Works	中国、台湾
	Global Wafers	日本、台湾、米国、イタリア、韓国
ドイツ	Siltronic	米国、ドイツ、シンガポール

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
*1: 信越化学工業、信越半導体などの信越グループ

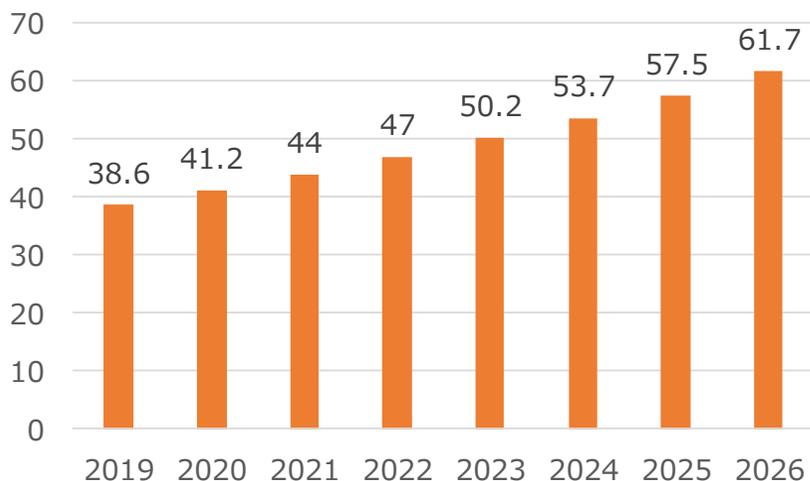
赤字は新設、もしくは新規掲載企業など

主要原料・主要部品の販売状況

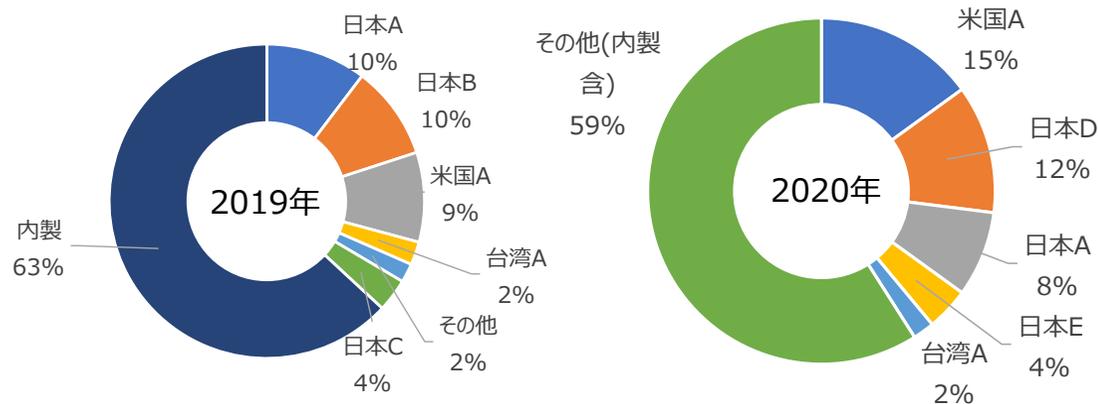
電子材料-フォトマスク

- フォトマスクの需要は、容量が増加し、電力管理、マイクロコントローラー、LED照明、MEMSで使用できるようになるにつれて、徐々に成長している。
- ディスプレイ用のフォトマスクは市場としては重要なセグメントであり続けると予想されている。
- 実際、2020年のフォトマスク市場規模は2019年から6.7%増で、電子材料の中では最も高く、今後も順調に増加すると予想されている。

億ドル 半導体フォトマスク世界市場規模推移



半導体フォトマスク売上高ベンダシェア



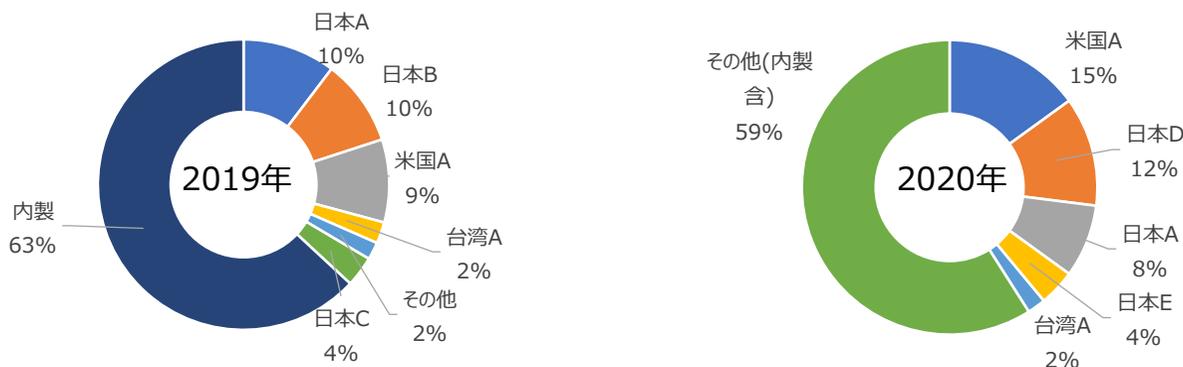
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-フォトマスク補足

- 2019年掲載上位企業のフォトマスクを扱う事業部門の2020年売上は好調で、2020年のフォトマスクのシェア状況については大きな変動はないと推察される。
- 2020年については、フォトマスク関連材料を含むシェアを示した。

半導体フォトマスク売上高ベンダシェア(再掲)



注) 2020年データは IndustryARC社の "Global Semiconductor Fabrication Material Market" のデータを基にNTTアドバンステクノロジー株式会社で作図

企業	フォトマスクを扱う事業部門	事業部門売上		2020年のフォトマスク売上に関連する記載
		2019年	2020年	
日本A	エレクトロニクス	1,781億円	1,837億円	半導体需要の高まりを受け、増収
日本B	エレクトロニクス	1,866億円	1,970億円	堅調、順調に推移
米国A	フォトマスクが主要事業	4.18億ドル (IC向け売上)	4.06億ドル (IC向け売上)	サプライチェーンの制約などの課題にも関わらず、IC市場でも順調に業績を伸ばした。

注)2020年はフォトレジスト関連材料も含めた数値

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-フォトマスク

- 凸版印刷は2018年に上海の施設で高度なフォトマスクを大量生産するための最先端機器に追加投資を行ったと発表している。
- DNPは米Photronicsと2018年から3年間の協業契約を行い、Photronicsの中国事業を支援している。Photronicsは、品質を損なうことなく、高速サイクルタイムと低コストに最適化されたAdvanced Binary Photomasksなどの製品を有する。

半導体フォトマスクにおけるサプライチェーン：各国・地域企業の生産地

国・地域	企業名	生産地
日本	DNP	日本、台湾、中国（Photronicsとの協業）、イタリア
	凸版印刷	日本、米国、台湾、中国、韓国、ドイツ、フランス
	HOYA	日本、中国（BOE合併：2021年10月発表）
	信越グループ	日本(フォトマスクブランクスを生産)
米国	Intel	米国
	Photronics	米国、イギリス、ドイツ、台湾、中国（DNP合併）
韓国	Samsung	韓国
台湾	TSMC	台湾
	Taiwan Mask	台湾
中国	无锡中微掩模（ZF Mask） ^{*1}	中国（無錫）

赤字は新設、もしくは新規掲載企業など

主要原料・主要部品の生産拠点の状況

電子材料-EUV露光用フォトマスク

- DNPは世界で初めてマルチ電子ビームマスク描画装置を導入し、次世代半導体用フォトマスクの生産性を向上させた。また、独自の製造プロセスを構築し、フォトマスク専業メーカーとしては初めて5nmプロセスノードEUVマスクを開発した。

EUV露光用フォトマスクにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	DNP	日本、IMEC（ベルギー）での開発にも参加
	凸版印刷 ^{*1}	欧州（ドイツ）、開発には米国からも参加
	HOYA ^{*2}	日本
米国	Photronics	米国（R&D拠点）

（出典）各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1: 2021年11月フォトマスク事業の会社分割による分社化を発表

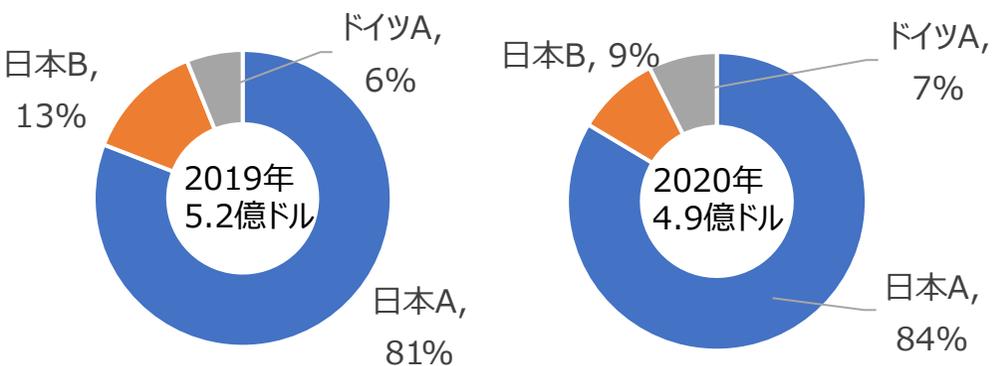
*2: EUV露光用フォトマスクの工場新設に関するニュースなし

主要原料・主要部品の生産拠点の状況

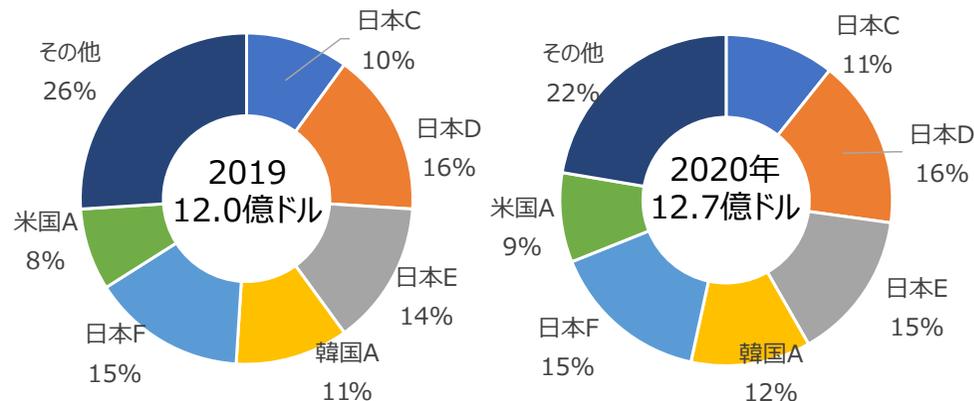
電子材料-フォトリソ関連装置および部材

- 電子ビームマスク描画装置*1は、2020年は9%ほど売上が落ちており、特に中国では約37%減ともっとも落ち込みが大きかった。
- マスクブランク市場はEUV用を中心に拡大している。また、AGCがEUV用ブランクの生産能力を増強*2、SKグループの投資会社(SK Japan Investment)がマスクブランクの日本企業を買収予定*3などの報道が確認された。

電子ビームマスク描画装置売上高シェア



マスクブランク売上高シェア



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1:電子線加工装置と走査型電子顕微鏡を応用し、電子銃から発せられた電子線を電子レンズやアパーチャー、デフレクタなどを通し、ステージを制御してマスクブランクスへ照射する装置。、EUV を使用する最先端プロセス向けにはマルチビーム描画装置が使用されるようになっている。(出典：世界半導体製造装置・試験／検査装置市場年鑑2021)

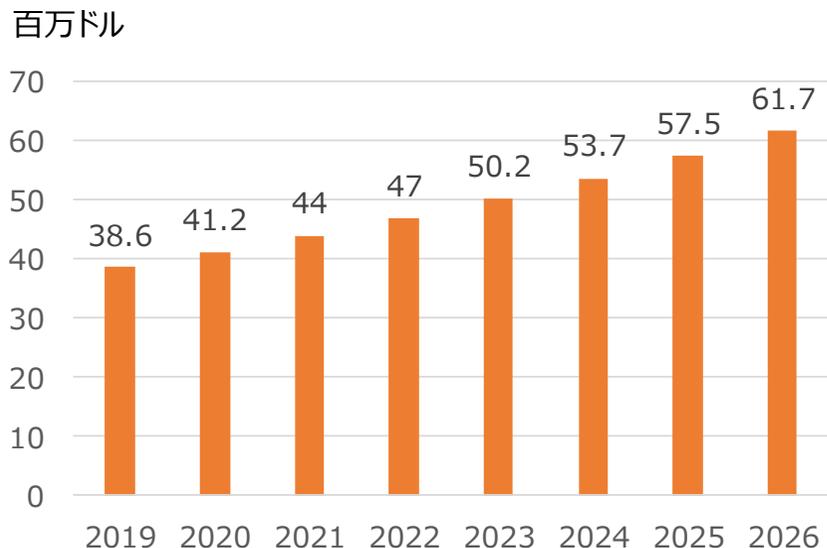
*2: <https://monoist.itmedia.co.jp/mn/articles/2202/17/news034.html>

*3: <https://news.mynavi.jp/techplus/article/20210527-1895245/>

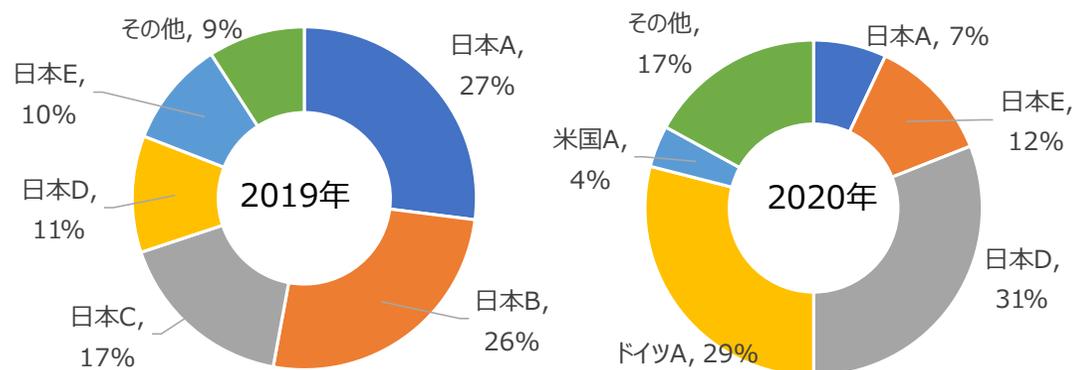
主要原料・主要部品の販売状況 電子材料-半導体露光用レジスト

- 2020年の半導体露光用レジスト市場規模は2019年から5.1%増で、今後も順調に増加すると予想されている。
- 半導体露光用レジストのシェアは日本企業が9割を超えている(2019年)。一方、後工程向けも含めたシェアでは、日本企業は5割程度となる(2020年)。

半導体露光用レジスト世界市場規模推移



半導体露光用レジストベンダ売上高シェア



注)2020年は後工程向けも含めた数値

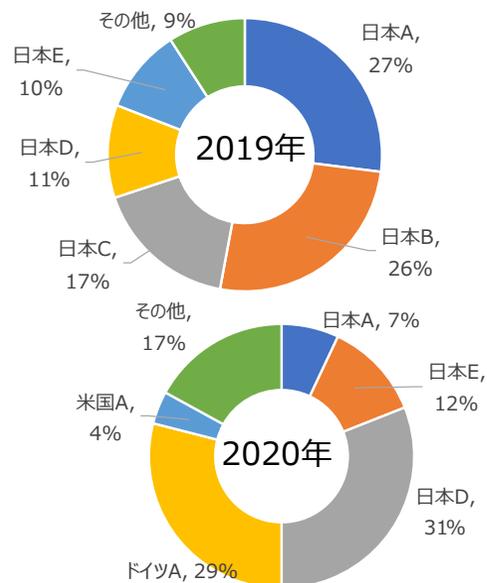
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-半導体露光用レジスト補足

- 2019年掲載各社の半導体露光用レジストを扱う事業部門の2020年売上は好調で、半導体露光用レジストのシェア状況については大きな変動はないと推察される。
- 2020年については、後工程向けも含めたシェアを示した。2020年のランキングに挙がっている企業は前工程向けレジストに関するノウハウを持っており、前工程向け半導体露光用レジストのシェア確保も狙える可能性がある。

半導体露光用レジストベンダ売上高シェア(再掲)



企業	半導体露光用レジストを扱う事業部門	2019年売上	2020年売上	2020年のレジスト売上に関連する記載
日本A	デジタルソリューション	1,448億円	1,514億円	主要顧客の先端デバイスが立ち上がるなど、最先端フォトレジストを中心に販売が堅調
日本B	エレクトロニクス機能材料	582億円	659億円	アジア地域において半導体用フォトレジストの売上増加
日本C	電子・機能材料	2,251億円	2,348億円	フォトレジスト製品は、ArFレジストやEUVレジストを中心に好調を持続
日本D	情報電子化学	4,049億円	4,318億円	需要の伸長に伴い出荷が増加
日本E	産業機材/電子材料/ファインケミカル	1,779億円	1,913億円	先端ロジック向けを中心に、幅広い製品群で販売が好調に推移、売上増加

注) 2020年データは IndustryARC社の“Global Semiconductor Fabrication Material Market”のデータを基にNTTアドバンステクノロジー株式会社で作図

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-半導体露光用レジスト

- 信越化学工業は、2020年に日本と台湾のフォトレジスト製造拠点到に300憶ドルを投資すると発表した。

半導体露光用レジストにおけるサプライチェーン：各国企業の生産地

企業国籍	企業名	生産地
日本	JSR	日本、米国、ベルギー
	東京応化工業	日本、韓国、米国、台湾、韓国
	信越グループ ^{*1}	日本、台湾
	住友化学	日本、韓国
	富士フイルム	日本、台湾、米国、中国、欧州(ベルギー)
米国	Dow / Dupont	日本、米国、韓国
ドイツ	Versum Materials ^{*2}	米国、韓国、台湾、日本、中国、ドイツ、フランス
韓国	Dongjin Semichem	韓国、台湾

赤字は新設、もしくは新規掲載企業など

(出典) NTTアドバンステクノロジー株式会社

*1: 信越化学工業、信越半導体など

*2: 2019年10月、Merckによって買収された。Merckの半導体材料生産拠点を表示。EUV露光用レジスト(ストリッパー)をどこで生産しているかは不明。

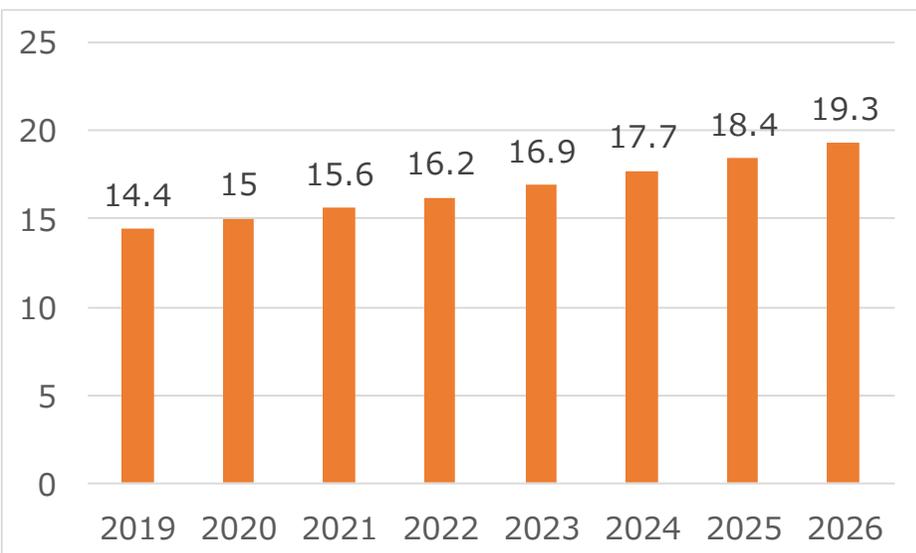
主要原料・主要部品の生産拠点の状況

電子材料-EUV露光用レジスト

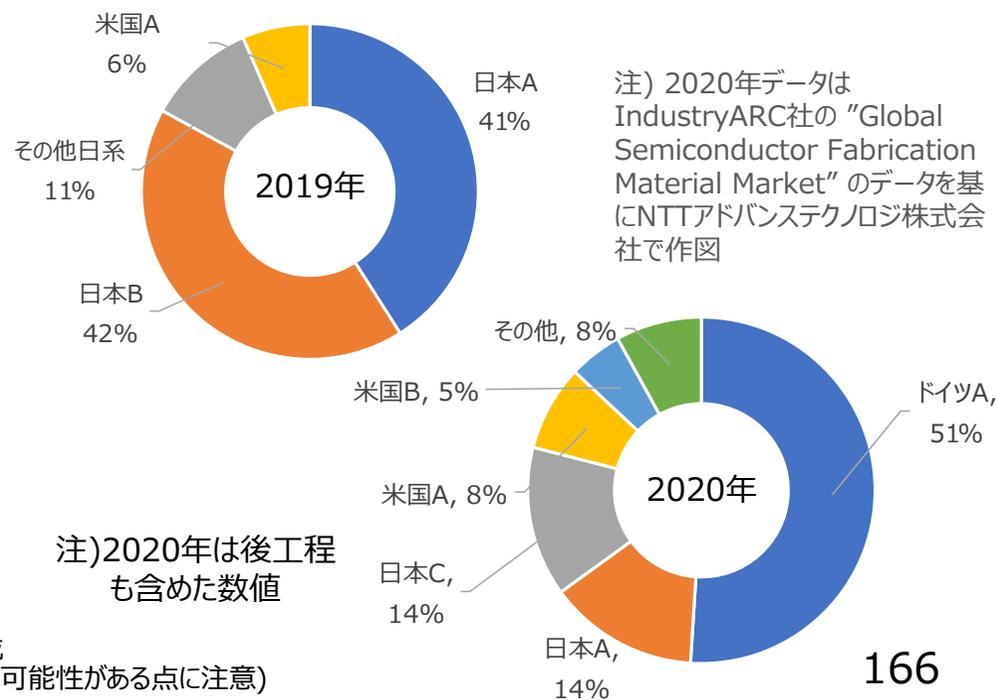
- 5Gデータ通信の普及により最先端プロセスで使うEUV露光用レジストの成長が見込まれる。
- 2020年のEUV露光用レジストの市場規模は2019年から4.1%増で、今後も順調に増加すると予想されている。
- EUV露光用レジストのシェアは日本企業が9割を超えている(2019年)。一方、後工程も含めたシェアでは、日本企業は3割程度となる(2020年)。

EUV露光用レジスト世界市場規模推移

百万ドル



EUV露光用レジストベンダ売上高シェア



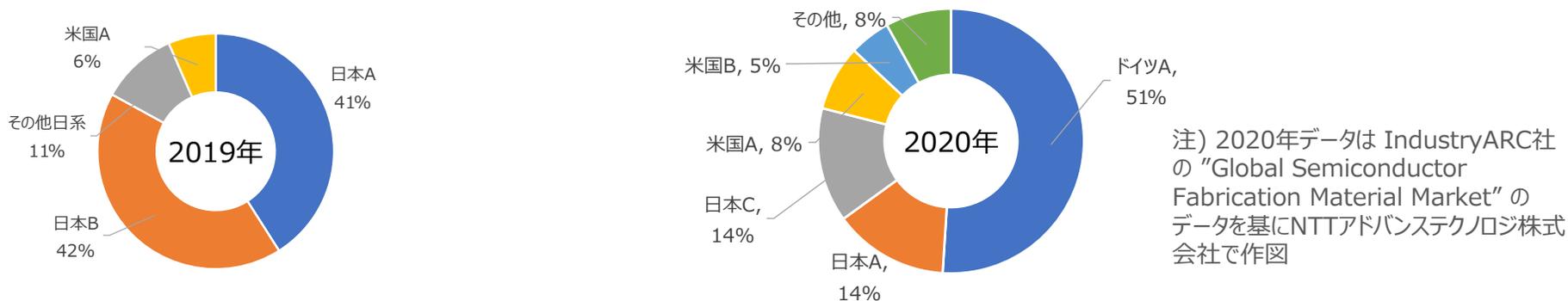
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-EUV露光用レジスト補足

- 2019年掲載各社のEUV露光用レジストを扱う事業部門の2020年売上は好調で、EUV露光用レジストのシェア状況については大きな変動はないと推察される。
- 2020年については、EUV露光用レジスト関連材料を含むシェアを示した。2020年のランキングに挙がっている企業はレジストに関するノウハウを持っており、EUV露光用レジスト自体のシェア確保も狙える可能性がある。

EUV露光用レジストベンダ売上高シェア(再掲)



企業	EUV露光用レジストを扱う事業部門	2019年売上	2020年売上	2020年のレジスト売上に関連する記載
日本A	デジタルソリューション	1,448億円	1,514億円	主要顧客の先端デバイスが立ち上がるなど、最先端フォトレジストを中心に販売が堅調
日本B	エレクトロニクス機能材料	582億円	659億円	アジア地域において半導体用フォトレジストの売上増加
米国A	Electronics & Imaging	35.5億ドル	38.1億ドル	レジストの売上が変化した等の記載無し

注)2020年はフォトレジスト関連材料も含めた数値

主要原料・主要部品の生産拠点の状況

電子材料-EUV露光用レジスト

- 住友化学は2021年ArF液浸やEUV露光用レジストの生産ラインを大阪に新設して2022年度上半期に稼働を開始、東京応化は韓国でのEUV露光用を含むフォトレジスト生産量を2021年末までに2倍以上に増やす、等各社増産体制を構築している模様。
- また、2021年9月、JSRは、SamsungにEUV露光用レジストを供給している米国Inpriaを買収したと発表しており、韓国での日本企業の影響力強化が懸念されている。

EUV露光用レジストにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	JSR	欧州（ベルギー） ^{*1} 、日本
	東京応化工業	日本、韓国(2020年より生産開始)
	信越グループ ^{*2}	日本、台湾(EUV露光用レジストを生産予定の台湾雲林工場の増設が2021年2月完成)
	住友化学	日本
	富士フィルム	日本
米国	Dow / Dupont	米国、韓国 ^{*3}
ドイツ	Versum Materials ^{*4}	米国、韓国、台湾、日本、中国、ドイツ、フランス

赤字は新設、もしくは新規掲載企業など

(出典) NTTアドバンステクノロジー株式会社

*1: imecとの合併会社EUV Resist Manufacturing & Qualification Center N.V. (EUV RMQC) で生産

*2: 信越化学工業、信越半導体などの信越グループ

*3: 韓国でのEUV露光用レジスト生産開始の報道はまだ見られない。一方、Samsungにはまだ供給していないという報道(2021年7月)は存在する。

*4: 2019年10月、Merckによって買収された。Merckの半導体材料生産拠点を表示。EUV露光用レジスト(ストリッパー)をどこで生産しているかは不明。

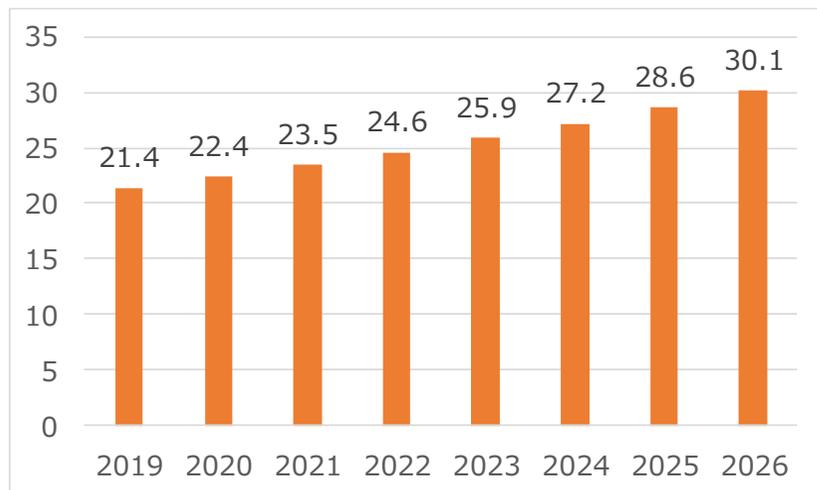
主要原料・主要部品の販売状況

電子材料-CMPスラリー・パッド

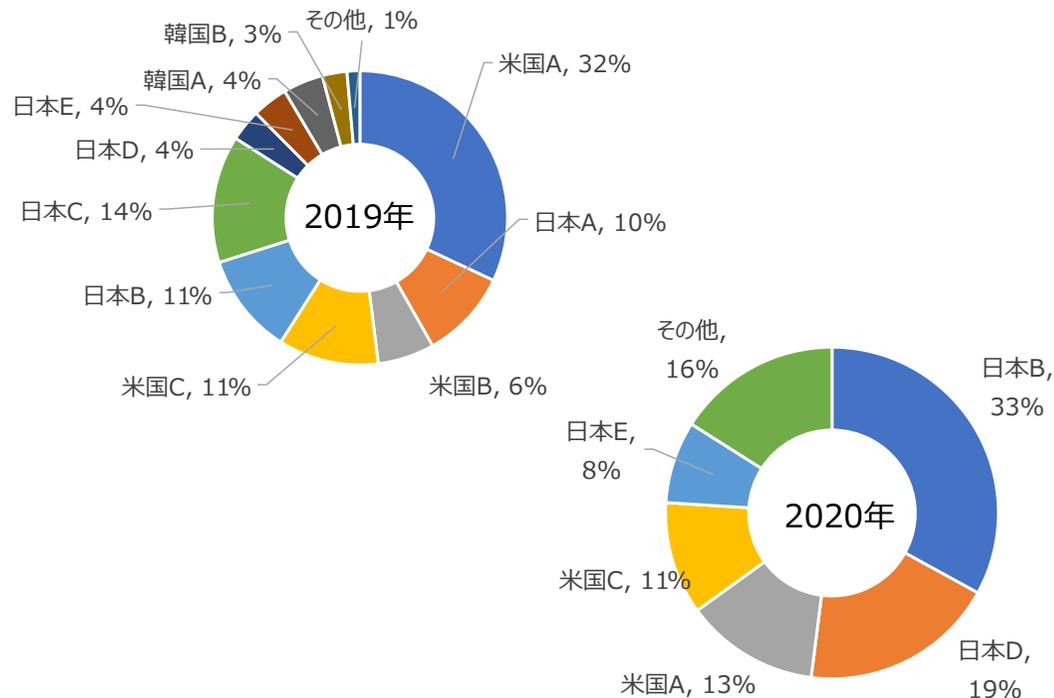
- サムスン電子のような企業だけでも、月に数万のCMPパッドを使用している。そのため、サムスン電子はCMPパッドの再利用技術を開発している。
- 2020年のCMPスラリー・パッド市場規模は2019年から4.8%増で、今後も順調に増加すると予想されている。

CMPスラリー・パッド世界市場規模推移

億ドル



CMPスラリー・パッドベンダ売上高シェア



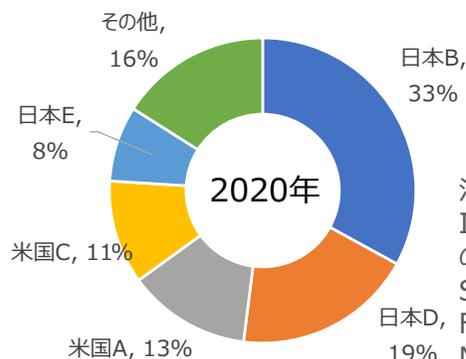
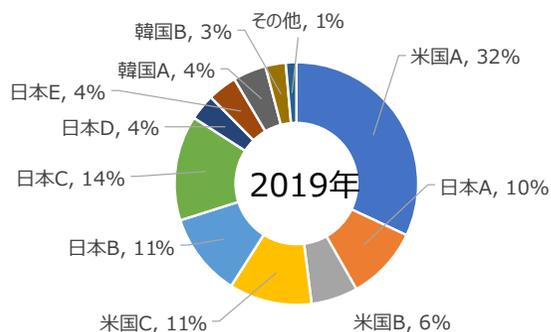
(出典) 各種情報源よりNTTアドバンステクノロジ株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-CMPスラリー・パッド補足

- 2019年シェア上位各社のCMPスラリーを扱う事業部門の2020年売上は好調で、CMPスラリーのシェア状況については大きな変動はないと推察される。
- 2020年については、CMPスラリー・パッドシェアを示した。2020年のランキングに挙がっている企業はCMPスラリーに関するノウハウを持っており、CMPスラリーのシェア確保も狙える可能性がある。

CMPスラリー・パッドベンダ売上高シェア(再掲)



注) 2020年データは IndustryARC社の "Global Semiconductor Fabrication Material Market" のデータを基に NTTアドバンステクノロジー株式会社で作図

企業	CMPスラリーを扱う事業部門	2019年売上	2020年売上	2020年のCMPスラリー売上に関連する記載
米国A	Electronic Materials	8.3億ドル	8.8億ドル	CMPスラリーの売上増がCMPパッドの売上減を補い増収
日本C	産業機材/電子材料/ファインケミカル	1,779億円	1,913億円	先端ロジック向けを中心に、幅広い製品群で販売が好調に推移、売上増加
日本B	機能材料	2,396億円	2,543億円	データセンター等の市場の伸長を受け、前年同期間を上回った
米国B	Electronics & Imaging	35.5億ドル	38.1億ドル	CMPスラリー・パッドの売上が変化した等の記載は無し
日本A	CMP向け製品セグメント	174億円	200億円	メモリ向けは一服したが、ロジック向けの需要は好調に推移、売上が大きく伸長

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-CMPスラリー・パッド

- 半導体デバイスの高性能化、高密度化、高集積化に伴い、CMPが適用されるプロセスは増加傾向にある。
- フジインコーポレーテッドなどは、顧客の製造・開発拠点到に近い、日本、米国、台湾に製造・開発拠点を設けている。また、昭和電工は台湾と韓国に総額200億円を投じCMPスラリーの製造能力を強化している。^{*1}

CMPスラリーにおけるサプライチェーン：各国企業の生産地

企業国籍	企業名	生産地
日本	フジインコーポレーテッド	日本、米国、台湾
	昭和電工マテリアルズ（旧日立化成）	日本、台湾、 韓国 ^{*2}
	富士フィルム ^{*3}	日本、米国、台湾、韓国
	JSR ^{*4}	日本
	AGC ^{*5}	日本、台湾、中国、欧州等
米国	Cabot Microelectronics(2020年10月よりCMC Materials) ^{*3}	日本、米国、台湾、韓国、欧州
	Dow Chemical/Dupont ^{*3}	米国、台湾

（出典）各種情報源によりNTTアドバンステクノロジー株式会社が作成

*1: <https://news.mynavi.jp/techplus/article/20201210-1576964/>

*2: 2021年10月稼働開始

*3: 2020年に工場新設や廃止のニュース無し

*4: 各拠点での生産製品種別は不明。

*5: HPのCMPスラリーのページ記載の拠点の一部。販売拠点も含む。

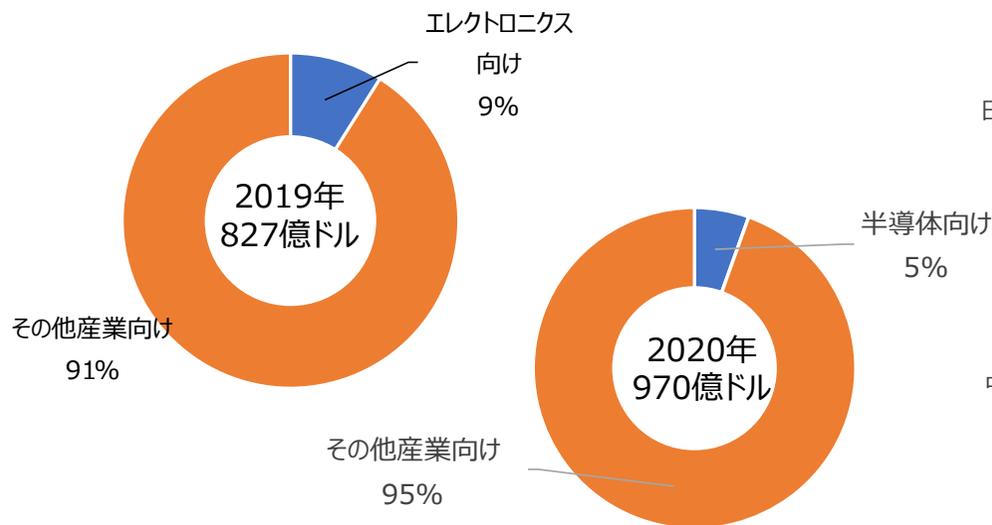
赤字は新設、もしくは新規掲載企業など

主要原料・主要部品の販売状況

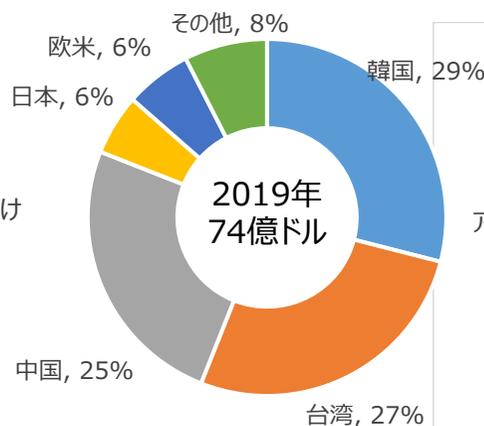
電子材料-産業ガス

- 産業ガス各社ではCOVID-19の影響からプラント操業率の低下など2020年第2四半期頃より業績に影響が出てきた。
- 2020年の半導体向け産業ガスの地域別売上シェアでは、TSMCやサムスン電子を含むアジア・太平洋が全体の50%を占めた。

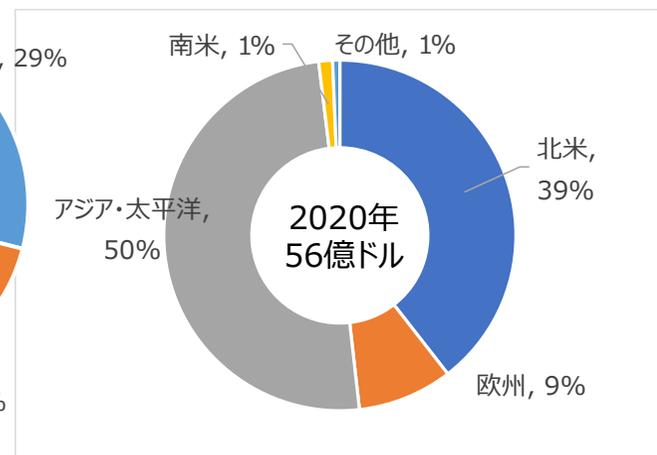
産業ガス用途別市場規模



産業ガス地域別シェア
(エレクトロニクス向け)



産業ガス地域別シェア
(半導体向け)



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

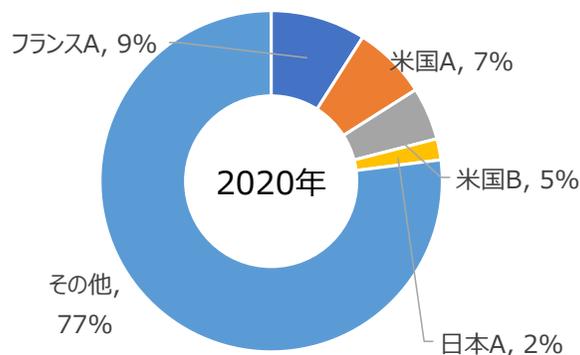
主要原料・主要部品の販売状況

電子材料-産業ガス

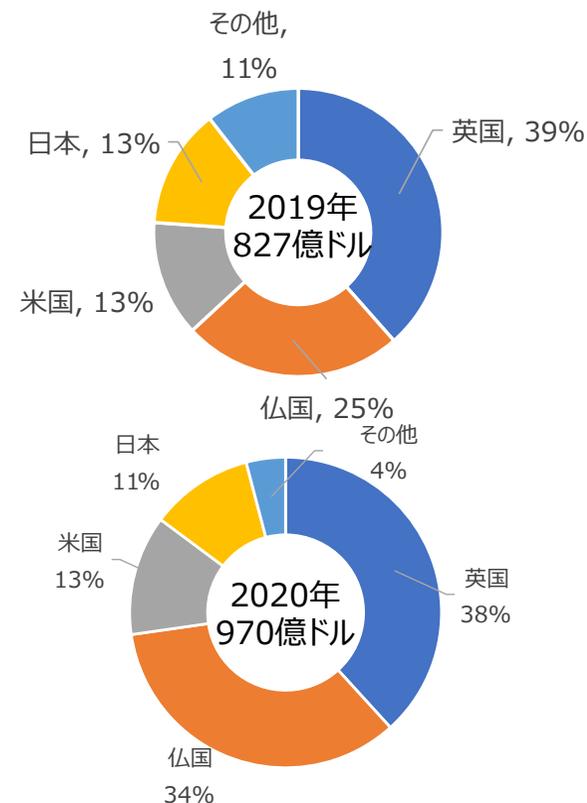
- 半導体の製造には、アルゴン、窒素、酸素、水素などの30を超える異なるガスが必要。
- 2020年の半導体向け産業ガスの市場規模は2019年から5.4%増で、今後も順調に増加すると予想されている。



産業ガスベンダシェア(半導体向)



産業ガスベンダ国・地域別シェア (全ての用途)



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-産業ガス

- 半導体用産業ガスは、主にベンダが半導体工場にてオンサイト供給を行っている。例えば大陽日酸は「トータルエレクトロニクス」という戦略に基づき、従来日本国内で製造していた特殊ガスをは中国や韓国でも製造するようになった。
- シェアのランキングには出ていないが、中国の718th、Huate、Nata、韓国のWonik、SK Materialsなども半導体向け産業ガスを販売している(⑤中韓装置材料メーカー調査参照)。

バルクガスにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
英国	Linde + Praxair	米国、欧州、アジア各国地域の半導体工場にてオンサイト供給
フランス	Air Liquide + Airgas	米国、欧州、アジア各国地域の半導体工場にてオンサイト供給
米国	Air Products	米国、欧州、アジア各国地域の半導体工場にてオンサイト供給
	Entegris ^{*1}	米国、韓国、マレーシア
日本	大陽日酸	米国、欧州、アジア各国地域の半導体工場にてオンサイト供給
中国	Yingde Gas Group	中国の半導体工場にてオンサイト供給

赤字は新設、もしくは新規掲載企業など

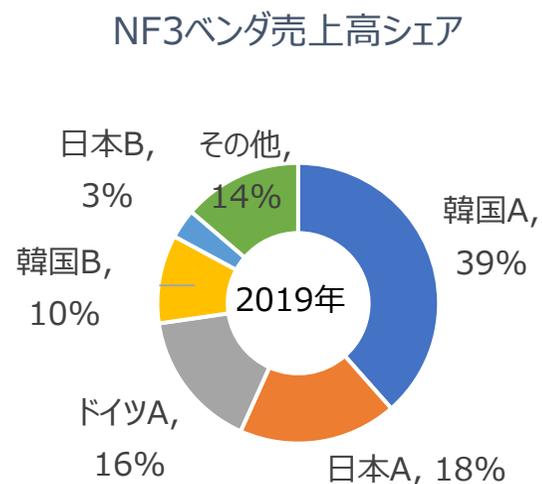
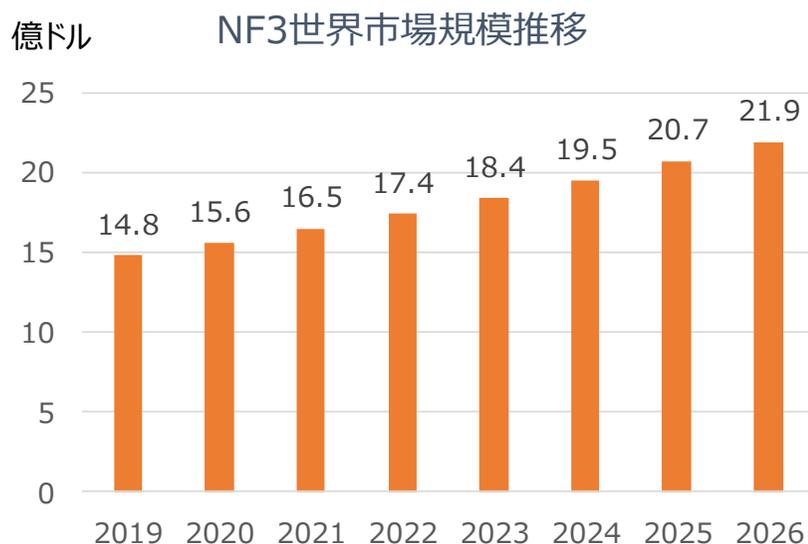
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1: 各拠点での生産製品種別は不明

主要原料・主要部品の販売状況

電子材料-特殊材料ガス（NF3）

- NF3は、製造プロセス中に電子部品のプロセスチャンバー壁、エッチング、および堆積アプリケーションから堆積物を除去するのに使用され、半導体ではシリコンウェーハの洗浄に使用される。
- 2020年のNF3市場規模は2019年から5.6%増で、今後も順調に増加すると予想されている。



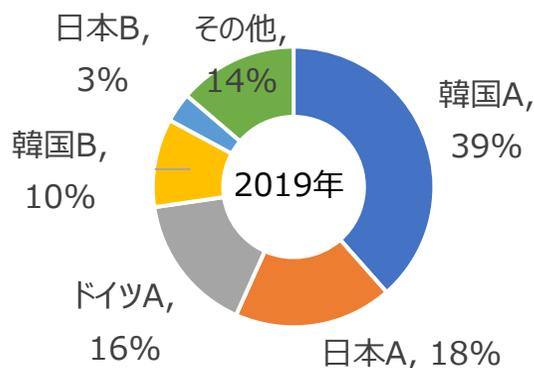
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-特殊材料ガス（NF3）

- 2019年シェア上位各社のNF3、もしくはNF3を扱う事業部門の2020年売上を以下に示す。NF3のシェア状況については大きな変動はないと推察される。

NF3売上高シェア(再掲)



企業	NF3を扱う事業部門	2019年売上	2020年売上	2020年のNF3売上に関連する記載
韓国A	NF3の売上	2,857億ウォン	2,865億ウォン	NF3の売上微増
日本A	精密化学品	400億円	403億円	販売数は増加したものの販売価格の低下により、前期に比べ減収
ドイツA	Performance Materials	25.7億ユーロ	33.8億ユーロ	NF3の売上に関する記載なし
韓国B	Chemical系子会社	18,125億ウォン	18,172億ウォン	NF3の売上に関する記載なし

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用)

主要原料・主要部品の生産拠点の状況

電子材料-特殊材料ガス（NF3）

- 2019年10月に Merck が Versum Materials を買収し、半導体電子材料を強化した。なお、米Entegris も Versum との統合を目指していた。今後も、材料企業の統合が進む可能性がある。
- 関東電化は2020年に中国に設立した子会社が現地に工場を建設中である。

NF3におけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	関東電化	日本、韓国、 中国（建設中）
	三井化学	日本、米国
韓国	SK Materials	韓国、中国
	Hyosung	韓国、中国
中国	PERIC	中国
ドイツ	Versum Materials ^{*1}	米国、韓国
英国	Linde	ドイツ、米国

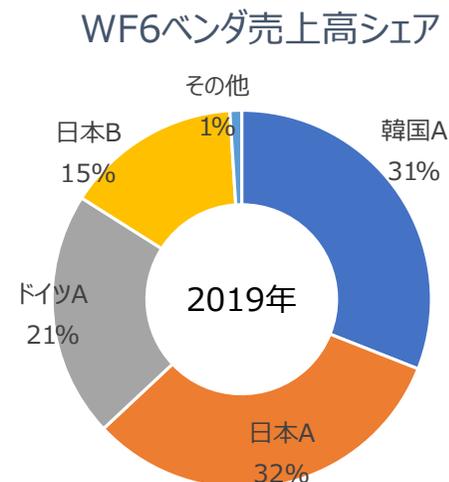
赤字は新設、もしくは新規掲載企業など

（出典）各種情報源よりNTTアドバンステクノロジー株式会社が作成
 *1: 2019年10月、Merckによって買収

主要原料・主要部品の販売状況

電子材料-特殊材料ガス(WF6)

- WF6は半導体をパターン化して金属接点とゲートを形成する際の化学蒸着プロセスで使用され、最高の出力を得るには、ガスの純度が99.999%を超えている必要がある。
- 3D NANDメモリの進化によりWF6の使用は増加しており、2020年のWF6市場規模は2019年から4.2%増で、今後も順調に増加すると予想されている。



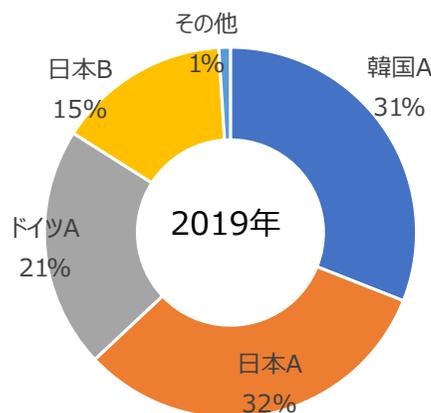
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-特殊材料ガス（WF6）

- 2019年シェア上位各社のWF6を扱う事業部門の2020年売上を以下に示す。WF6のシェア状況については大きな変動はないと推察される。

WF6売上高シェア(再掲)



企業	WF6を扱う事業部門	2019年売上	2020年売上	2020年のWF6売上に関連する記載
韓国A	WF6の売上	927億ウォン	1,146億ウォン	WF6の売上増加
日本A	精密化学品	400億円	403億円	販売価格は低下したものの販売数量の増加により、前期に比べ増収
ドイツA	Performance Materials	25.7億ユーロ	33.8億ユーロ	WF6の売上に関する記載なし(技術開発進展の記載あり)
日本B	ファインケミカル	782億円	783億円	堅調な半導体需要により半導体用途の特殊ガス関連製品の出荷が増加

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用)

主要原料・主要部品の生産拠点の状況

電子材料-特殊材料ガス（WF6）

- セントラル硝子は、中国浙江博瑞中硝科技有限公司との合併で浙江博瑞電子科技有限公司を設立し、中国でのWF6の開発と生産を行っている。
- 関東電化は韓国忠清南道にあるグループ会社関東電化ファインプロダクツ韓国株式会社でフッ素化合物の製造販売を行っており、2020年に中国に設立した子会社が現地に工場を建設中である。
- このように、日本企業が半導体企業の近くで生産する動きが見られる。

WF6におけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	関東電化	日本、韓国、中国（建設中）
	セントラル硝子	日本、中国（合併）
韓国	SK Materials	韓国
	Foosung	韓国
ドイツ	Versum Materials ^{*1}	米国
中国	Chengdu Taiyu Industrial Gases	中国
	浙江博瑞中硝科技有限公司	中国（セントラル硝子との合併企業）

赤字は新設、もしくは新規掲載企業など

（出典）各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1: 2019年10月、Merckによって買収

主要原料・主要部品の販売状況

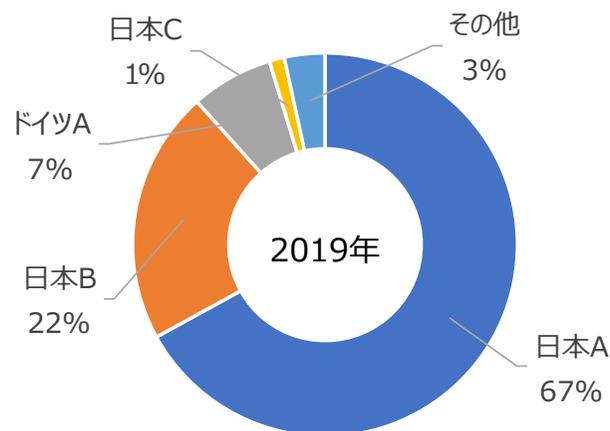
電子材料-特殊材料ガス（クリーニングガス、LPCVD）

- 三フッ化塩素（ ClF_3 ）は効果的な酸化剤として機能し、難燃性のアスベストとして機能する耐火性の発火物により、半導体業界で表面の酸化物を洗浄するのに使用される。
- 2020年のクリーニングガス市場規模は2019年から5.8%増で、今後も順調に増加すると予想されている。

クリーニングガス(ClF_3 ,20% F_2 / N_2)
世界市場規模推移



クリーニングガスベンダ売上高シェア

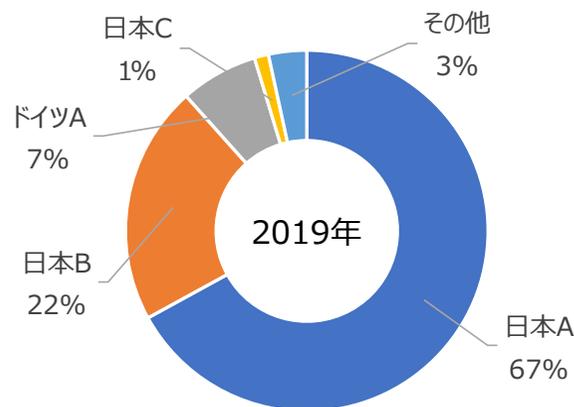


(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況 電子材料-特殊材料ガス（クリーニングガス）

- 2019年シェア上位各社のクリーニングガスを扱う事業部門の2020年売上を以下に示す。クリーニングガスのシェア状況については大きな変動はないと推察される。

クリーニングガス売上高シェア(再掲)



企業	クリーニングガスを扱う事業部門	2019年売上	2020年売上	2020年のクリーニングガス売上に関連する記載
日本A	ファインケミカル	782億円	783億円	堅調な半導体需要により半導体用途の特殊ガス関連製品の出荷が増加
日本B	精密化学品	400億円	403億円	クリーニングガスの売上に関する記載なし
ドイツA	Performance Materials	25.7億ユーロ	33.8億ユーロ	クリーニングガスの売上に関する記載なし

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用)

主要原料・主要部品の生産拠点の状況 電子材料-特殊材料ガス（クリーニングガス、LPCVD）

- SK Materialsは昭和電工との合併会社(SK Showa Denko)を設立し、クリーニングガスとしてCH₃Fを生産開始した。

クリーニングガスにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	セントラル硝子	日本
	関東電化	日本、韓国、中国（建設中）
韓国	SK Materials	韓国（昭和電工との合併会社）
ドイツ	Versum Materials ^{*1}	米国

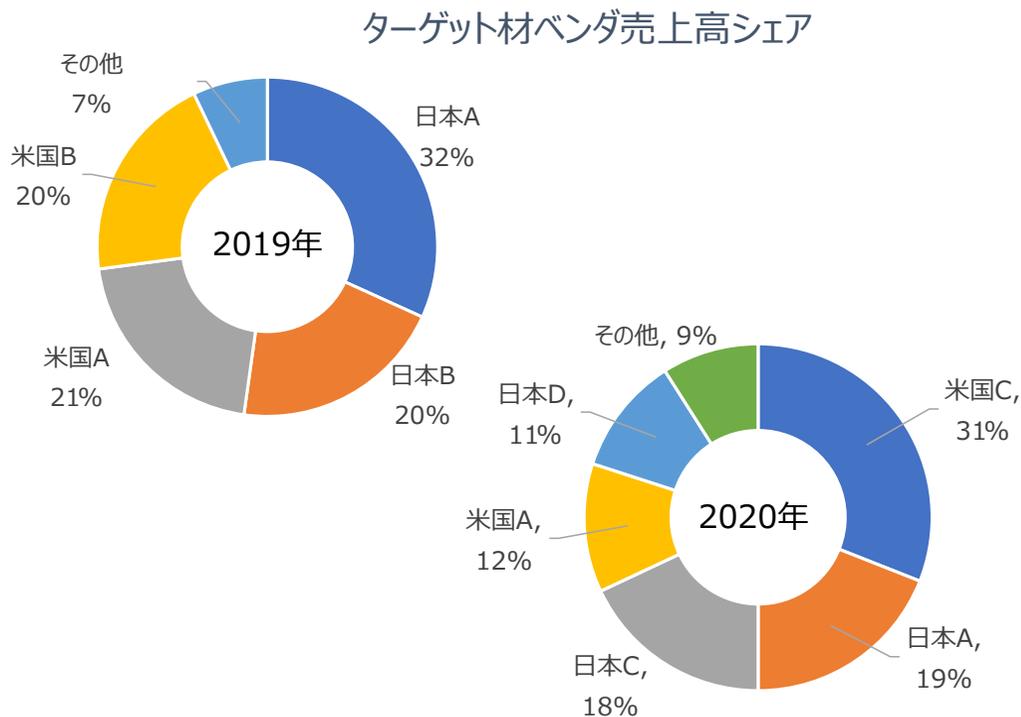
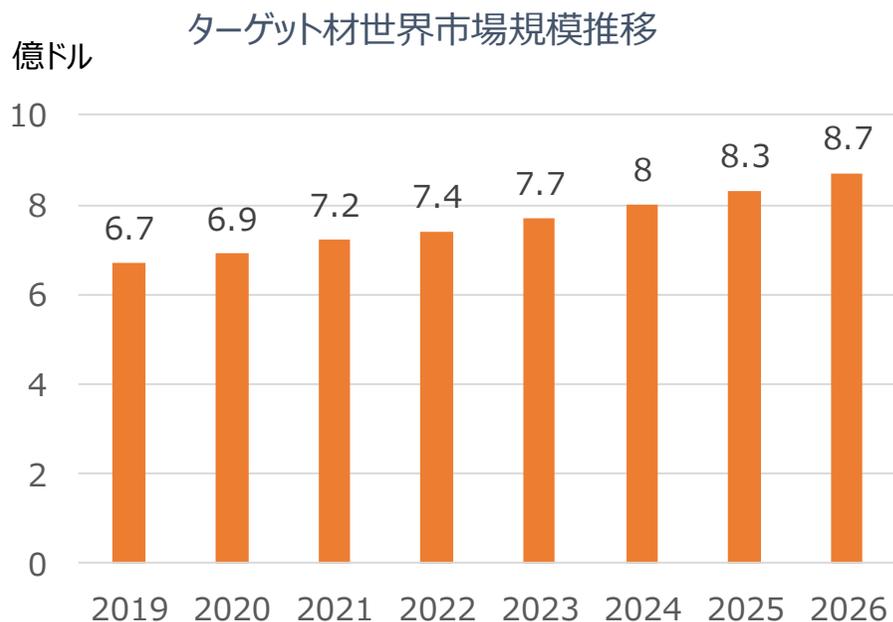
赤字は新設、もしくは新規掲載企業など

（出典）各種情報源よりNTTアドバンステクノロジー株式会社が作成
*1: 2019年10月、Merckによって買収

主要原料・主要部品の販売状況

電子材料-ターゲット材

- 半導体のスパッタリングターゲットは、主に導電層、バリア層、および金属グリッドの製造に使用され、主にアルミニウム、チタン、銅、タンタル、およびその他の金属が原料となる。
- 2020年のターゲット材の市場規模は2019年から3.6%増で、今後も順調に増加すると予想されている。



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

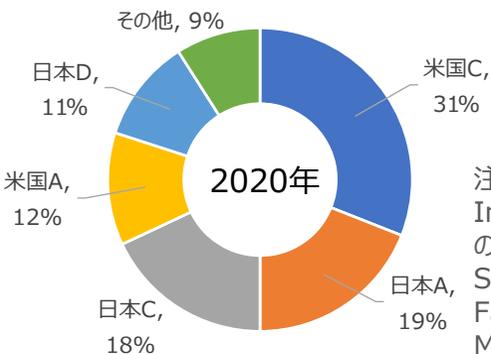
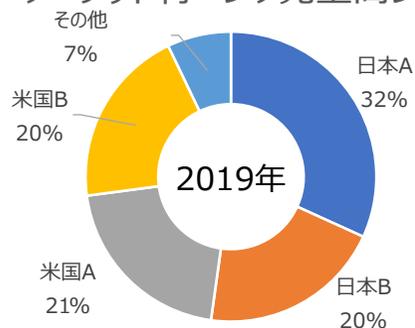
注)2020年はターゲット材関連材料を含むシェア

主要原料・主要部品の生産拠点の状況

電子材料-ターゲット材補足

- 2019年掲載各社は売上を落としている企業が多かった。しかしその額はわずかなので、ターゲット材のシェア状況については大きな変動はないと推察される。
- 2020年については、ターゲット材関連材料を含むシェアを示した。2020年のランキングに挙がっている企業はターゲット材に関するノウハウを持っており、ターゲット材自体のシェア確保も狙える可能性がある。

ターゲット材ベンダ売上高シェア(再掲)



注) 2020年データは IndustryARC社の "Global Semiconductor Fabrication Material Market" のデータを基に NTTアドバンステクノロジー株式会社で作図

企業	ターゲット材を扱う事業部門	2019年売上	2020年売上	2020年のターゲット材売上に関する記載
日本A	金属	10,897億円	10,044億円	高機能IT分野での需要増加を主要因に機能材料・薄膜材料は堅調に増販
日本B	機能商品	1,976億円	1,933億円	機能商品事業を中心に販売数量が計画を下回る
米国A	Advanced Materials	28.0億ドル	26.6億ドル	機能性材料は、特殊品の販売数量は増加したが、フッ素製品などの販売数量が減少
米国B		282億ドル	272億ドル	ターゲット材の売上に関する記載無し

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
(2019年データは 令和2年度報告書より引用
2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-ターゲット材

- 2020年JX金属は、ターゲット材の生産能力を30%引き上げることを決定した。
- 東ソーは将来の半導体素子の微細化に対応した次世代配線用の有機金属化合物、製造プロセスを刷新可能なGaNスパッタリングターゲットを開発している。

ターゲット材におけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	JX金属	日本、米国、韓国、台湾
	東ソー	日本、米国、韓国
	日立金属 ^{*1}	日本、米国、中国、タイ、インド、フィリピン
	京セラ	日本
	Honeywell Electronic Materials ^{*2}	日本、米国、韓国、ドイツ
米国	Praxair ^{*2}	米国、中国、ドイツ
	Materion ^{*3}	米国、ドイツ、アイルランド、シンガポール、台湾、中国
中国	KFMI（江丰电子） ^{*4}	中国、マレーシア ^{*5}

（出典）各種情報源によりNTTアドバンステクノロジー株式会社が作成

*1: パワーエレクトロニクス事業としての生産地

*2: 2020年に工場新設や廃止のニュース無し

*3: ターゲット材を含むAdvanced Materialsという分類の生産地

*4: 三菱ケミカルがKFMIに軽金属事業を売却(2020年11月) https://www.m-chemical.co.jp/news/2020/1210225_7469.html

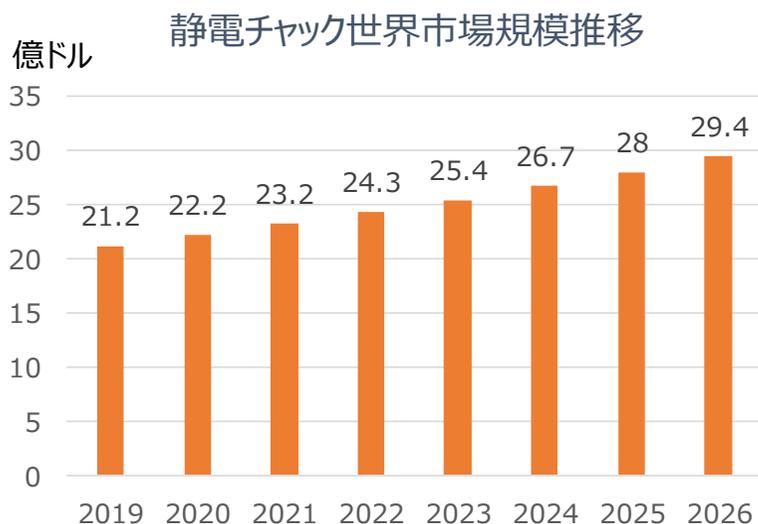
*5: マレーシアでは半導体用ではなく太陽電池用スパッタリングターゲット生産

赤字は新設、もしくは新規掲載企業など

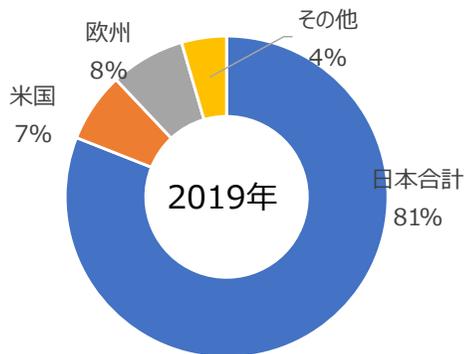
主要原料・主要部品の販売状況

電子材料-静電チャック

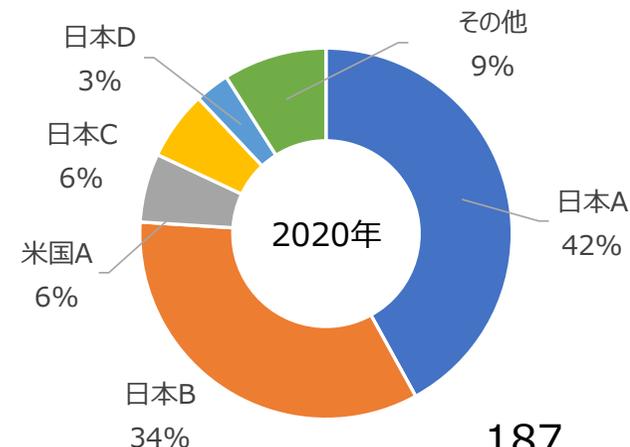
- 2020年にSMICが12インチのシリコンウェーハの製造を開始することを発表した。シリコンウェーハのサイズが大きいほど、保持できるマイクロチップの数が増えるため、中国では8～12インチの静電チャックの成長率が高くなると予想されている。
- 実際、2020年の静電チャックの市場規模は2019年から4.6%増で、今後も順調に増加すると予想されている。



静電チャック国・地域別シェア



静電チャック企業別シェア



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料-静電チャック

- 2020年のシェアデータでランクインしたEntegrisのSCEM(SPECIALTY CHEMICALS AND ENGINEERED MATERIALS)セグメントは米国、韓国、マレーシアに製造拠点を有する。

静電チャックにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	新光電気工業 ^{*1}	日本
	日本ガイシ	日本
	日本特殊陶業(NGK)	日本
	京セラ	日本、米国
	TOTO	日本
	住友大阪セメント	日本
米国	II - IV / M Cubed ^{*2}	米国
	Entegris ^{*3}	米国、韓国、マレーシア
ドイツ	Protec	ドイツ

赤字は新設、もしくは新規掲載企業など

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1: 2020年に工場新設や廃止のニュース無し

*2: II - VIはM CubedをM&Aで吸収

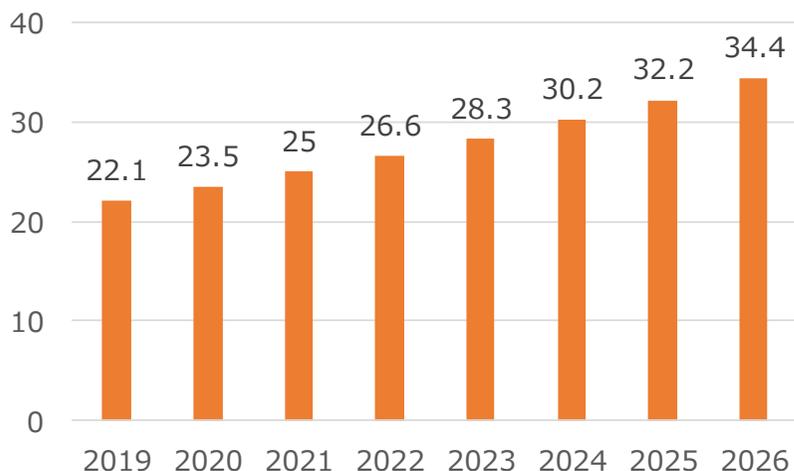
*3: 各拠点での生産製品種別は不明

主要原料・主要部品の販売状況

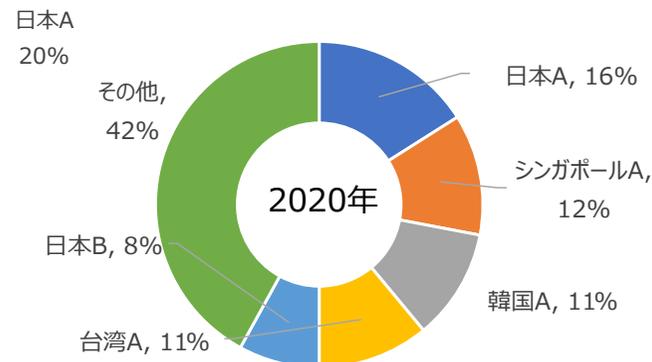
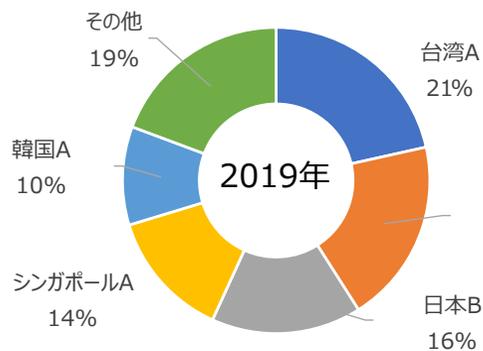
電子材料-IC用リードフレーム

- IC用リードフレームは、半導体チップを固定して外部と接続するものであり、リードフレームが薄いほど、ICの半導体製造における密着性が高まり、信頼性が高まる。
- 2020年のIC用リードフレームの市場規模は2019年から6.3%増で、今後も順調に増加すると予想されている。
- 2020年のシェア Top 5企業は2019年と変わっていない。

億ドル IC用リードフレーム世界市場規模推移



IC用リードフレームベンダ売上高シェア



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

主要原料・主要部品の生産拠点の状況

電子材料- IC用リードフレーム

- 日中台の他、マレーシアに工場をもつ企業が多い点が目立つ。
- 2021年10月に大日本印刷が半導体QFNタイプのパッケージ用リードフレームを開発*4したと発表するなど技術競争が激しい。

IC用リードフレームにおけるサプライチェーン：各国・地域企業の生産地

国・地域	企業名	生産地
台湾	Chang Wah Electronics	中国、マレーシア、台湾
日本	三井ハイテック	日本、マレーシア、シンガポール、台湾、中国
	新光電気工業	日本、マレーシア、中国
	大日本印刷	日本
シンガポール	ASM Pacific*1	中国、シンガポール、マレーシア
韓国	Haesung DS*1	韓国、中国、日本

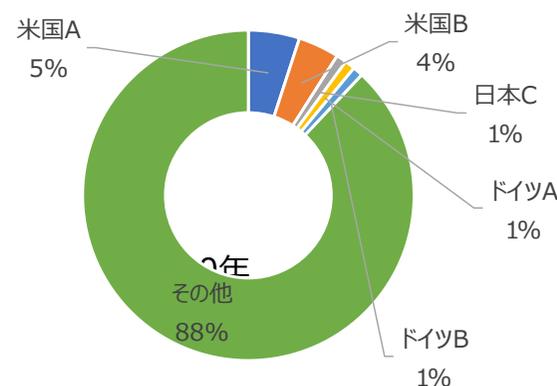
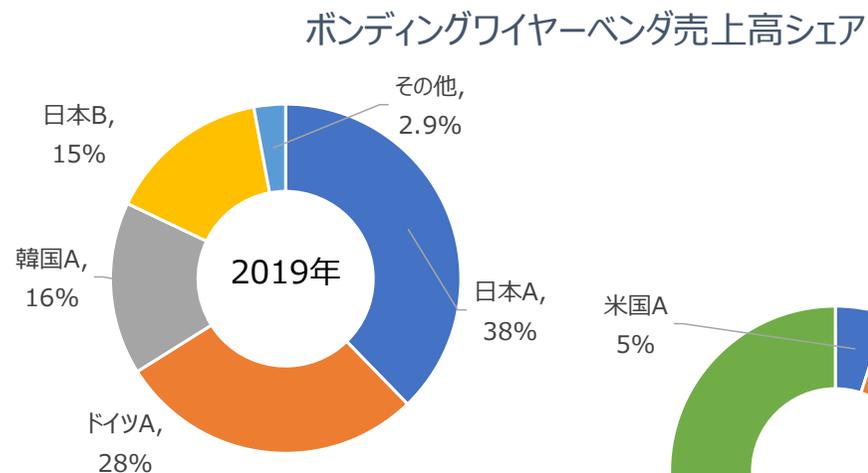
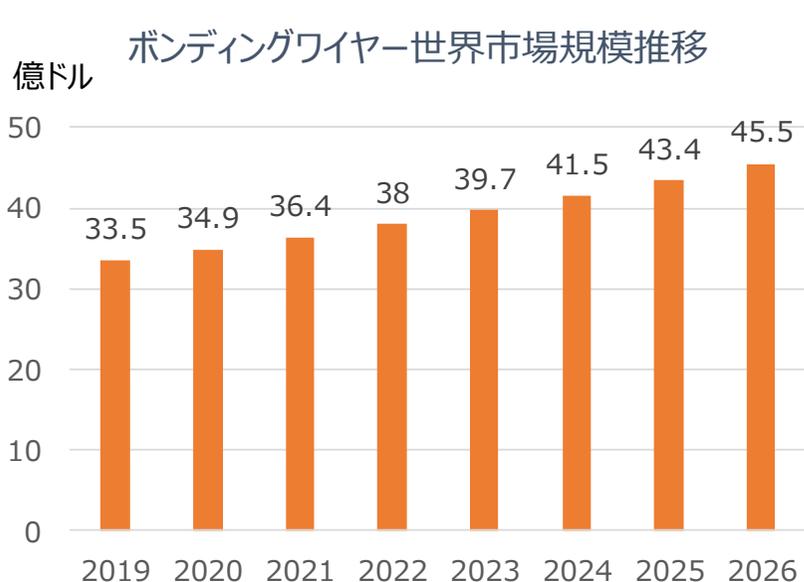
(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成

*1: 各拠点での生産製品種別は不明

主要原料・主要部品の販売状況

電子材料-ボンディングワイヤー

- ボンディングワイヤーはICとの衝撃の強い機械的接触を必要とし、特に壊れやすい次世代のガリウムベースのチップとパッケージ内の余分なスペースでスクラップを生成することが多く、設計者が家電製品の需要を満たす洗練されたデバイスを作成することを困難にする。
- 2020年のボンディングワイヤーの市場規模は2019年から4.3%増で、今後も順調に増加すると予想されている。



(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成
 (2019年データは 令和2年度報告書より引用
 2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意)

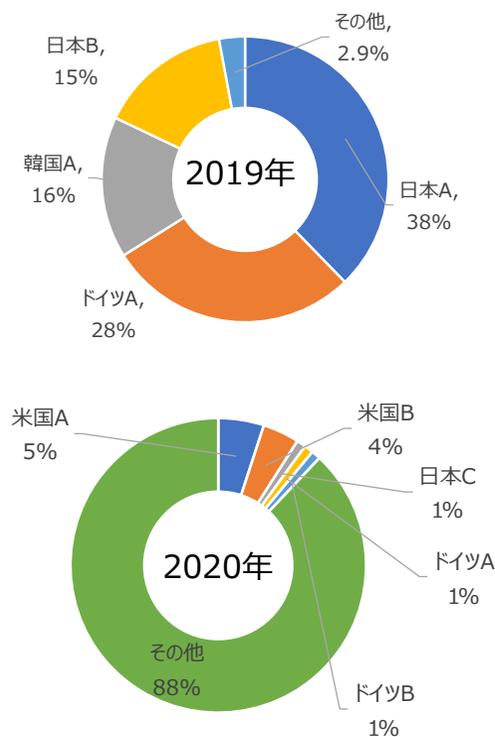
注)2020年はボンディングワイヤー関連材料を含むシェア

主要原料・主要部品の生産拠点の状況

電子材料-ボンディングワイヤー補足

- 2019年掲載各社のボンディングワイヤーを扱う事業部門の2020年売上は好調で、ボンディングワイヤーのシェア状況については大きな変動はないと推察される。^{*1}
- 2020年については、ボンディングワイヤー関連材料を含むシェアを示した。

ボンディングワイヤーベンダ売上高シェア(再掲)



注) 2020年データは IndustryARC社の "Global Semiconductor Fabrication Material Market" のデータを基に NTTアドバンステクノロジー株式会社で作図

企業	ボンディングワイヤーを扱う事業部門	2019年売上	2020年売上	2020年のボンディングワイヤー売上に関連する記載
日本A	ボンディングワイヤーの製造販売が中心事業	254億円	283億円	有価証券報告書なし
ドイツA		224億ユーロ	315億ユーロ	Annual Reportなし
韓国A	ボンディングワイヤー	4,432億ウォン	5,830億ウォン	韓国内の売上は減ったが、国外での売上が増加
日本B	機能材料	560億円	600億円	年度を通じて堅調に推移、スマートフォン向け材料の販売が回復に転じて増加

(出典) 各種情報源よりNTTアドバンステクノロジー株式会社が作成

(2019年データは 令和2年度報告書より引用)

2019年と2020年の情報源が異なるため、集計方法などが異なる可能性がある点に注意

*1: Annual Reportに「ボンディングワイヤー技術は、微細化、高強度化、高信頼性化が核心であり、短期間では習得しにくい技術であり、産業部門の特性上、新規参入が困難」と記載している企業あり。

主要原料・主要部品の生産拠点の状況

電子材料- ボンディングワイヤー

- 田中電子工業は台湾の高雄市に高性能パラジウム被覆Cu(銅)ボンディングワイヤーの工場を建設中で、工場稼働は2022年上期を予定している。また、2022年1月には中国の杭州市にパワー半導体用Al(アルミニウム)ボンディングワイヤーの第2工場を新設し、2025年までに生産能力を約3倍に増やす計画と発表。金に代わってコストの安いワイヤーの製造が増える可能性がある。

ボンディングワイヤーにおけるサプライチェーン：各国企業の生産地

国・地域	企業名	生産地
日本	田中電子工業	日本、中国、台湾、マレーシア、シンガポール
	日鉄マイクロメタル	日本、フィリピン、中国
	タツタ電線	日本、マレーシア
	新光電気 ^{*1}	日本、韓国、中国、マレーシア
アメリカ	Honeywell ^{*2}	米国、韓国、中国、欧州等
	Materion ^{*3}	米国、ドイツ、アイルランド、シンガポール、台湾、中国
ドイツ	Heraeus ^{*4}	ドイツ、中国、マレーシア、シンガポール
	Henkel ^{*5}	米国、ドイツ、韓国、中国等
韓国	MK Electron	韓国、中国
	Heesung Metal	韓国 (LT Metal, 田中貴金属工業との合弁)、中国

(出典) 各種情報源によりNTTアドバンステクノロジー株式会社株式会社が作成

*1: 各拠点での生産製品種別は不明。

*2: HP記載のロケーションより抜粋。生産地及び生産製品種別に該当するかは不明。

*3: Advanced Materialsという分類での生産地。

*4: 2020年に工場新設や廃止のニュース無し

*5: HP記載のロケーションより抜粋。販売拠点も含む。

赤字は新設、もしくは新規掲載企業など

1-2-9. まとめ

総括 (1/2)

- 電子機器・半導体関連市場はCOVID-19による巣ごもり需要、DX需要により堅調に伸びている。
- Foundry市場では、TSMCが最先端プロセスで先行し、シェアを拡大している。また、SMICは中国市場の旺盛な需要を引き受けシェアを拡大している。
- IDM市場では、スーパーサイクルの好影響を受けたメモリメーカーであるSK Hynix、Samsung、Micronがシェアを拡大している。ビッグデータの普及によるデータセンタでのDRAM需要が大きく、DRAM主体のSK Hynixの伸びが大きい。
- Fabレス市場では、Intelが10nmプロセス開発に失敗しているあいだにAMDがシェアを拡大している。また、NVIDIAの主力製品であるGPUがPCゲームでの利用、データセンタでの利用、暗号通貨のマイニングでの利用が伸び、シェアを拡大している。
- OSAT市場およびEMS市場では、中国・台湾のメーカーが市場を支配しており、M&Aなどにより積極的にシェア確保に動いている。
- 半導体製造装置市場は、半導体関連市場の動きにあわせて2018年までのスーパーサイクルの反動で2019年に減少したものの、2020年はCOVID-19による巣ごもり需要、DX需要により堅調に伸びている。
- 半導体材料市場は、各材料の2019年から2020年の市場の伸びについては、ターゲット材が3.6%で最も低く、フォトマスクが6.7%で最も高かったが、全ての材料が3-7%の範囲で伸びており、市場の伸びが目立つ材料は無かった。

総括 (2/2)

- 半導体材料メーカーに関しては、買収(MerckによるVersum Materialsの買収)や日本企業と海外企業の合併(DNPとPhotronics、セントラル硝子と中国ガス企業、昭和電工とSK Materials、LT Metalと田中貴金属工業などの合併による合併)による企業間の技術移転・連携が見られた。一方、国から買収が阻止されるケース(ドイツ政府によるGlobalwafersのSiltronic買収不成立)もあり、各国・地域が材料メーカーの重要性を認識しつつあるのではないかと推定される。

展望

- COVID-19による消費の落ち込みからの回復やニューノーマルな生活様式の定着により、自動車、PC、サーバ、ゲーム機等の需要が増加しているが、半導体の供給が追いつかず世界的な半導体不足が顕在化している。
- 半導体不足解消に向け、Foundry・IDM・Fabレス・OSAT・EMS・半導体製造装置・半導体材料の各市場は今後も拡大すると思われる。ただし、どこかで需要と供給の逆転が起きるのはこれまでのシリコンサイクルの示すところである。
- OSAT・EMSでは中国メーカーのシェアが大きく、今後も中国の存在感が増すと考えられる。また、半導体製造装置市場では中国市場のシェアがトップとなっており、今後中国での半導体製造が増えると思われる。

1-3. 主要生産国におけるマイクロエレクトロニクスに係る生産に対する公的支援の実績と効果の調査

1-3-1. 調査の目的と背景	199
1-3-1-1. 調査方法	201
1-3-2. 地域毎の実績の調査	204
1-3-2-1. 地域毎の調査結果	205
1-3-2-2. 地域毎・年毎の実績	214
1-3-3. 調査結果分析	220
(参考) 主要企業の調査	228
1-3-4. 公的支援に関する最近の話題	234
1-3-4-1. 地域毎の公的支援に関する最近の話題	235
1-3-5. まとめ	245

1-3-1. 調査の目的と背景

1-3-1. 調査の目的と背景

主要生産国である、米・中・韓・台・独の5カ国・地域について、マイクロエレクトロニクスに係る生産に対する公的支援の実績と効果を調査する。

【背景】

- 最近になって、各国が、経済安全保障の観点から重要な半導体の生産基盤を囲い込むため、巨大な公的支援策を打ち出してきている。
- 一方、以前より中国では、国家ICファンドなどにより、半導体設計、製造、材料、装置への大規模な公的支援が行われてきた。それに対して、米国、ドイツ、韓国、台湾については個別プロジェクト・案件など個別対応が多かった。

【目的】

- これら各国・地域の支援の実績を調査し、定量的に比較し、分析を行うことで、これまでの公的支援の効果を検証する。
- また、最近の公的支援に関する話題を整理することで、今後の方向性を展望する。

1-3-1-1. 調査方法

調査方法

- 各国・地域の半導体産業への公的支援情報のうち、金額と期間が明確な情報を、オープンソースデータ*1から収集した。
- 公的支援には補助金・ファンド、税制優遇、研究開発支援、インフラ優遇などがある。*2
- 上記のうち、オープンソースデータ*1から、各国・地域の半導体産業に関連する公的支援情報のうち、金額と期間が明確な情報を収集した。
- 具体的には、以下のような条件で情報を収集した。
 - 実績値が分かる場合は実績値の金額、成立・可決した予算金額のみが分かる場合はそれらの金額を収集した。
 - 半導体に限定できない(例えば「××技術」など)場合で、半導体の割合が推測できる場合は推測値を収集した。
 - 例えば「半導体、自動運転、バイオに2兆ウォンの投資」の場合、半導体は1/3の6,667億ウォンと推定した。
- 代表的な企業については、各企業のAnnual Reportから情報を収集した。
- 調査結果の分析に関しては、オープンソースデータから各国・地域の半導体製造能力、半導体製造装置売上額、Fabless企業の売上額などのデータを収集し、公的支援金額との関係を考察した。

*1: 各国・地域の半導体産業に関するニュースサイトを中心に情報を収集した。

*2: 令和2年度重要技術管理体制強化事業(https://www.meti.go.jp/meti_lib/report/2020FY/000167.pdf)

公的支援の実績の具体的な収集方法

- 各国・地域のニュースサイトで検索し、金額と期間が明確な情報を収集した。
- 韓国と台湾を中心に、主要企業のAnnual Reportからも情報を収集した。

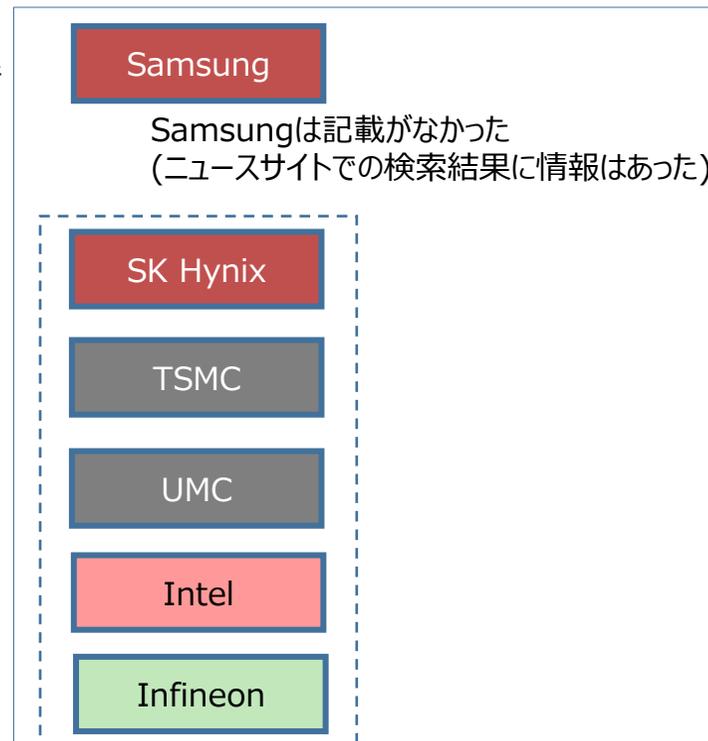
ニュースサイト*1

日本語のサイト	"半導体" "補助金"
米国のサイト	"semiconductor" "subsidy"/"government grant"
中国のサイト	"半导体" "补贴"
韓国のサイト	"반도체" "보조금"
台湾のサイト	"半導體" "補貼"
ドイツ・欧州のサイト	"Halbleiter" "Subventionen"

特に、韓国と台湾の情報が少なかったため、主要半導体企業のAnnual Reportも確認



Annual Report



これらの企業は記載があったので、Annual Reportより収集
(ニュースサイトでの検索結果と重複すると判断した場合はどちらか一方のみとした)

※中国については、国全体での公的支援情報がニュースとして取り上げられているため、個別企業の情報がニュースサイトでの検索結果に含まれていると判断した

*1: 次頁以降のリスト

1-3-2. 地域毎の実績の調査

1-3-2-1. 地域毎の調査結果

実績の調査：調査結果(米国 1/2)

- 収集した米国の主な公的支援金額情報を以下に掲載する。

年	年数	金額 (億)	単位	概要
2013	1	0.4	米ドル	【R&D】DARPA による半導体研究センター6カ所の設立
2013	5	0.3	米ドル	【R&D】米国の半導体研究協会と NIST によるナノエレクトロニクス研究イニシアチブへの支援
2013	1	0.3	米ドル	【R&D】ARPA-E による広バンドギャップ半導体のための革新的な材料、装置設計、装置製造プロセスを追求するプロジェクト
2013	5	1.7	米ドル	【R&D】ニューヨーク州による炭化ケイ素を使用したウェーハ製造 (SUNYPolyプログラム)
2013	1	1.3	米ドル	【Fab】インテルが獲得した資産、プラント、装置に関する政府補助金
2014	1	1.0	米ドル	【Fab】インテルが獲得した資産、プラント、装置に関する政府補助金
2014	1	48.0	米ドル	【R&D】ニューヨーク州によるインテル、TSMC、サムスンなど5社に対する次世代半導体開発への補助金
2014	5	3.5	米ドル	【Fab】DoEによるノースカロライナ州にパワーエレクトロニクスの製造拠点を設置するための補助金
2015	1	1.1	米ドル	【Fab】インテルが獲得した資産、プラント、装置に関する政府補助金
2017	1	0.04	米ドル	【R&D】DoDによるAIサーバ向けGPUの研究開発助成金
2017	1	0.005	米ドル	【R&D】DoDによるMil-spec memoryの開発助成金
2017	1	0.02	米ドル	【Fab】ワシントン州の半導体製造、半導体材料の製造に使用されるガスおよび化学物質の販売および使用、製造施設の建設税の税率優遇
2017	1	2.6	米ドル	DoEによるインテルなどへの助成金
2017	1	0.7	米ドル	バージニア州によるMicronへの助成金

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

実績の調査：調査結果(米国 2/2)

- 収集した米国の主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2018	1	2.2	米ドル	【R&D】DARPAによる半導体技術開発プロジェクト(ERI)
2018	1	1.6	米ドル	【Fab】インテルが獲得した資産、プラント、装置に関する政府補助金
2019	1	1.7	米ドル	【R&D】国防総省による宇宙放射線に耐えられるチップの生産など、生産能力の向上と技術の向上支援
2019	1	0.1	米ドル	【Fab】テキサス州によるTIへの設備投資助成金(Texas Enterprise Fund)
2019	1	0.4	米ドル	【Fab】ニューヨーク州によるFab10買収の助成金・減税
2019	1	6.1	米ドル	【Fab】ニューヨーク州によるCreeIncへの200mm電源およびRFウェーハ製造工場助成金
2020	1	1.4	米ドル	国防総省によるインテルへの補助金(SHIP(State-of-the-art Heterogeneous Integration Prototype)第2フェーズ)

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

実績の調査：調査結果(中国)

- 収集した中国の主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2013	1	300.0	元	省・自治区・直轄市・市政府などの地方政府によるIC関連投資ファンド
2014	5	1,387.0	元	中国政府による中国IC産業投資ファンド(ビッグファンド)第一期
2014	1	300.0	元	北京市による集積回路産業の発展のための株式投資ファンド
2015	1	500.0	元	【Fab】上海市による先進的なIC生産ラインの建設と設備・材料産業への支援(基金)
2015	1	318.0	元	省・自治区・直轄市・市政府などの地方政府によるIC関連投資ファンド
2016	1	2,080.0	元	省・自治区・直轄市・市政府などの地方政府によるIC関連投資ファンド
2017	1	900.0	元	省・自治区・直轄市・市政府などの地方政府によるIC関連投資ファンド
2019	1	0.2	元	【R&D】重慶市による製造業の研究開発に対する補助金
2019	1	500.0	元	深セン市政府によるICサブファンド
2019	1	16.0	元	中国政府によるChangXin Memoryへのシンジケートローン
2019	1	19.0	元	半導体上場企業20社への中国政府の補助金
2020	5	2,041.5	元	中国政府による中国IC産業投資ファンド(ビッグファンド)第二期
2020	1	22.5	米ドル	中国政府系ファンドからFoundry最大手の中芯国際集成电路製造(SMIC)への出資
2020	1	600.0	円	中国政府系ファンドから清華紫光集團傘下の半導体ファブレスIC設計会社UNISOC Communicationsへの出資
2020	1	149.1	元	中国政府による半導体企業119社への補助金

(出典) NTTアドバンステクノロジ株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

実績の調査：調査結果(韓国 1/2)

- 収集した韓国の主な公的支援金額情報を以下に掲載する。

年	年数	金額 (億)	単位	概要
2013	1	76.4	ウォン	【Fab】SK Hynixに対する建物と設備に関する補助金
2014	1	71.7	ウォン	【Fab】SK Hynixに対する建物と設備に関する補助金
2014	1	1.8	ウォン	【R&D】韓国政府による、サムスン電子への研究インフラ補強などの補助金
2015	1	66.2	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金
2016	1	58.1	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金
2017	1	58.3	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金
2017	1	25.0	ウォン	サムスン電子に対する韓国政府からの補助金
2018	1	195.1	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金
2018	1	116.0	ウォン	サムスン電子に対する韓国政府からの補助金
2018	1	111.0	ウォン	韓国政府によるフォトレジストやエッチングガス企業への補助金
2019	1	147.0	ウォン	サムスン電子に対する韓国政府からの補助金
2019	1	10.9	ウォン	【R&D】韓国政府による、TESへの半導体プロセスや装置に関する研究開発補助金
2019	1	184.8	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

実績の調査：調査結果(韓国 2/2)

- 収集した韓国の主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2020	1	131.0	ウォン	【R&D】LG電子に対する韓国政府からの研究開発補助金
2020	1	1,100.0	ウォン	韓国政府による素材・部品・装備ベンチャーファンド
2020	1	800.0	ウォン	韓国政府による次世代産業技術政策ファンド
2020	1	82.0	ウォン	サムスン電子に対する韓国政府からの補助金
2020	3	6,666.7	ウォン	韓国政府による5兆ウォン規模の投資(「素材・部品・装備2.0」)。このうち、半導体、未来自動車(自動運転車やEV)、バイオの3大産業に2兆ウォン規模の投資を行う。
2020	1	172.8	ウォン	【Fab】SKハイニックスに対する建物と設備に関する補助金
2020	1	150.0	ウォン	【Fab】韓国政府による国内へ戻ってくる企業の工場への補助金
2020	1	23.2	ウォン	【R&D】韓国政府によるTESへの半導体プロセスや装置に関する研究開発補助金

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

実績の調査：調査結果(台湾)

- 収集した台湾の関係機関による主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2014	1	0.2	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2015	1	2.5	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2016	1	15.4	台湾ドル	【Fab】TSMCのプラント、設備、土地使用に対する補助金
2016	2	0.2	台湾ドル	【R&D】A+企業イノベーション研究開発プログラム
2016	1	95.7	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2017	1	67.6	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2017	1	28.1	台湾ドル	【Fab】TSMCのプラント、設備、土地使用に対する補助金
2018	4	40.0	台湾ドル	【R&D】AI端末で使用する半導体の製造とICチップシステムに関する研究への補助(半導体射月計画(半導体エッジプロジェクト))
2018	1	71.3	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2018	1	151.0	台湾ドル	台湾当局によるChipMOST Technologyへの融資
2019	1	6.2	台湾ドル	【Fab】UMCに対する資産取得に対する補助金
2019	1	34.2	台湾ドル	【Fab】TSMCのプラント、設備、土地使用に対する補助金
2020	1	10.7	台湾ドル	【Fab】TSMCのプラント、設備、土地使用に対する補助金
2020	2	76.9	台湾ドル	台湾当局による半導体分野への補助金
2020	1	2.4	台湾ドル	【Fab】UMCに対する資産取得に対する補助金

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

実績の調査：調査結果(ドイツ、欧州 1/2)

- 収集したドイツ(ドイツを含む欧州)の主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2013	1	0.3	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2013	7	100.0	ユーロ	【R&D】ECが主導した半導体開発プロジェクト(Horizon2020)。70%を参加各国政府・地方自治体が、残りの30%をEU(欧州連合)が拠出。ドイツのドレスデンなどが研究開発拠点。
2014	1	0.4	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2015	1	0.4	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2015	3	0.6	ユーロ	バイエルン州によるパワー半導体などの産業クラスターへの補助金
2016	1	0.5	ユーロ	ベルギー・フランダース政府のimecへの補助金
2016	1	0.3	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2017	1	1.1	ユーロ	ベルギー政府のimecへの補助金
2017	1	2.0	ユーロ	ザクセン州によるボッシュへの補助金
2017	1	3.0	ユーロ	ドイツ財務省によるドイツのチップ部門への補助金
2017	1	17.0	ユーロ	【Fab】ドイツ政府によるドレスデンの拠点設立などのマイクロエレクトロニクスへの補助金
2017	1	0.4	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2017	1	0.3	ユーロ	【R&D】英国のISCF(産業戦略チャレンジ基金)。
2017	1	2.0	ユーロ	【Fab】シュヴァーベン自動車サプライヤーによる、ドレスデンの新チップ工場建設のための補助金

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

実績の調査：調査結果(ドイツ、欧州 2/2)

- 収集したドイツ(ドイツを含む欧州)の主な公的支援金額情報を以下に掲載する。

年	年数	金額(億)	単位	概要
2018	1	0.4	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2019	3	0.7	ユーロ	【R&D】8カ国・地域、43のパートナーによる高出力密度と高エネルギー効率のパワー半導体の研究開発のための欧州の共同プロジェクト(Power2Power)
2019	3	0.3	ユーロ	【R&D】ASMLなどによる3nm半導体技術の統合に関するパイロット研究(Pin3sプロジェクト)に対するEUの補助金。
2019	1	0.005	ユーロ	【R&D】英国政府機関からParagrafに対する、半導体中のインジウムをグラフェンに置き換えるための研究の補助金。
2019	1	0.6	ユーロ	【Fab】Infineonの製造施設建設のための補助金
2020	1	0.4	ユーロ	【Fab】Infineonの製造施設建設のための補助金

【Fab】は、生産能力拡充目的が明確な案件、
【R&D】は、研究開発目的が明確な案件

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

1-3-2-2. 地域毎・年毎の実績

実績の調査：調査結果

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額を示す。
- 中国の公的支援金額は他国・地域に比べると巨大である。
- 中国以外では、欧州や米国の補助金額が大きい。

順位	公的支援金額*1			
	研究開発含		生産能力拡充、その他*2	
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)
1	中国	110,525.4	中国	110,062.3
2	欧州	15,419.8	欧州	3,433.9
3	米国	8,475.6	ドイツ	3,247.0
4	ドイツ	3,247.0	米国	1,972.8
5	台湾	1,981.8	台湾	1,873.8
6	韓国	488.8	韓国	466.7

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

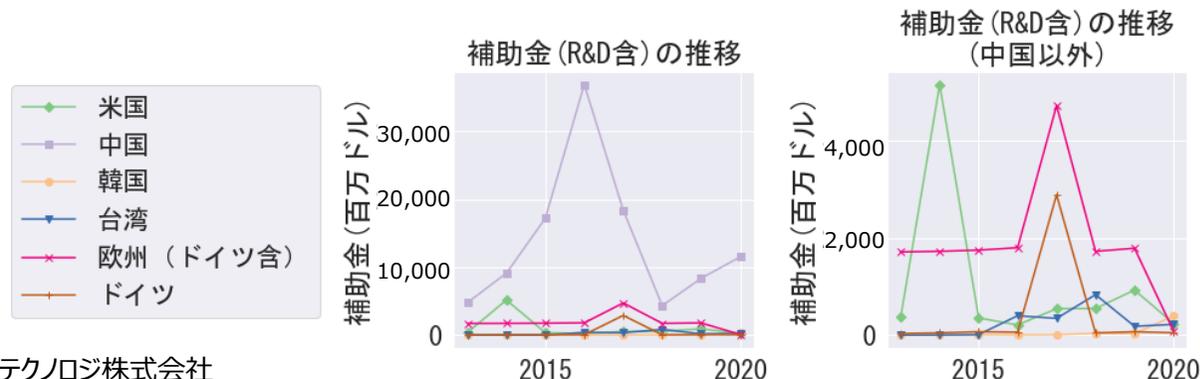
*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

実績の調査：調査結果

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額(研究開発目的を含む)を集計した結果を示す。

年	米国	中国	韓国	台湾	ドイツ	欧州（ドイツ含）
2013	371.3	4,770.9	6.3		31.7	1,709.3
2014	5,151.3	9,097.3	6.1	0.8	44.6	1,722.2
2015	352.3	17,176.2	5.5	9.1	68.5	1,746.1
2016	208.5	36,858.8	4.8	397.7	60.3	1,796.6
2017	542.3	18,363.1	6.9	342.6	2,883.3	4,719.7
2018	546.0	4,326.4	36.7	831.2	43.4	1,721.0
2019	925.1	8,346.6	29.0	180.2	68.1	1,785.8
2020	240.0	11,494.0	392.2	220.2	47.0	86.4

単位は百万米ドル*1



(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

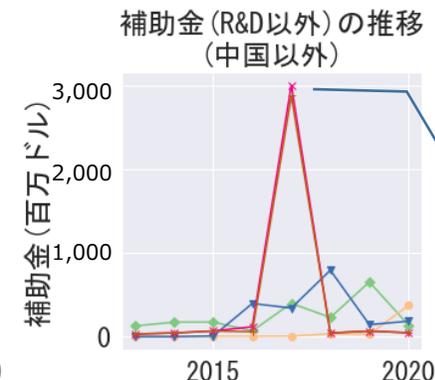
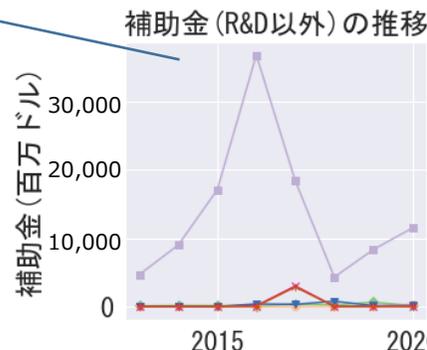
実績の調査：調査結果

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額(主に生産支援、研究開発目的を除く)を集計した結果を示す。

年	米国	中国	韓国	台湾	ドイツ	欧州（ドイツ含）
2013	129.0	4,678.9	6.3		31.7	31.7
2014	174.0	9,005.3	5.9	0.8	44.6	44.6
2015	175.0	17,084.2	5.5	9.1	68.5	68.5
2016	70.0	36,766.8	4.8	397.4	60.3	119.0
2017	399.7	18,363.1	6.9	342.3	2,883.3	3,011.6
2018	230.0	4,326.4	34.9	795.4	43.4	43.4
2019	655.1	8,343.5	27.4	144.5	68.1	68.1
2020	140.0	11,494.0	373.8	184.4	47.0	47.0

単位は百万米ドル*1

中国政府によるIC産業投資ファンド(ビッグファンド)
(2014-2018)+地方政府によるIC関連投資ファンド



ドイツ政府によるドレスデンの拠点設立などのマイクロエレクトロニクスへの補助金

(出典) NTTアドバンステクノロジー株式会社

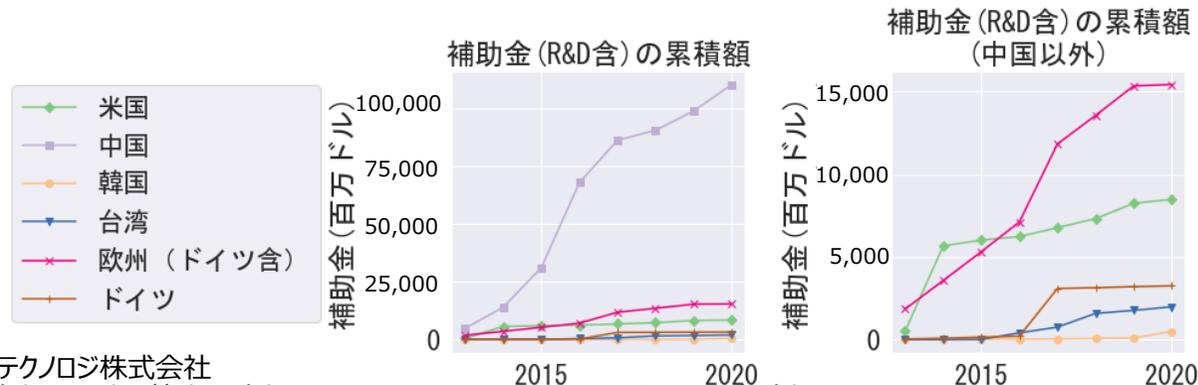
*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

実績の調査：調査結果

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額(研究開発目的を含む)の2013年からの累積額を示す。

年	米国	中国	韓国	台湾	ドイツ	欧州（ドイツ含）
2013	510.1	4,862.9	7.6		31.7	1,842.0
2014	5,661.4	13,960.2	13.7	0.8	76.3	3,564.2
2015	6,013.7	31,136.4	19.2	9.9	144.8	5,310.3
2016	6,222.2	67,995.2	24.0	407.6	205.1	7,106.9
2017	6,764.5	86,358.3	30.8	750.3	3,088.4	11,826.6
2018	7,310.5	90,684.8	67.5	1,581.4	3,131.9	13,547.6
2019	8,235.6	99,031.4	96.5	1,761.7	3,200.0	15,333.4
2020	8,475.6	110,525.4	488.8	1,981.8	3,247.0	15,419.8

単位は百万米ドル*1



(出典) NTTアドバンステクノロジー株式会社

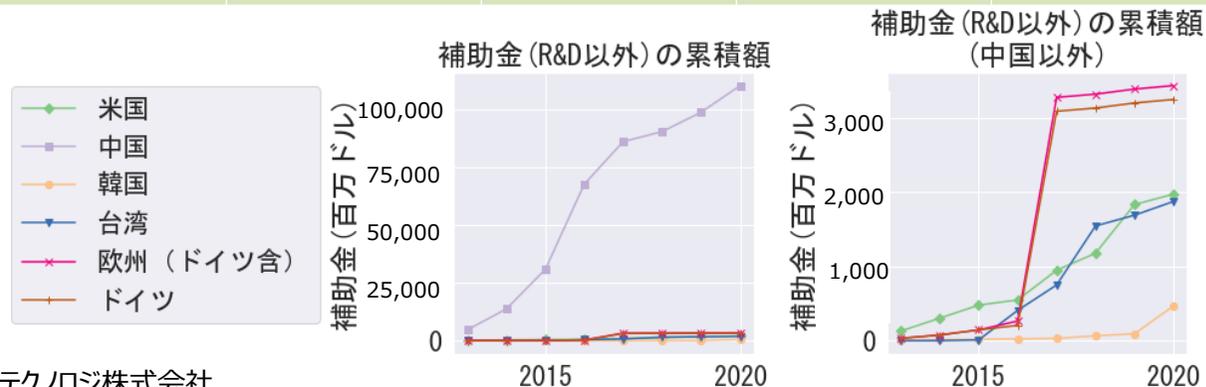
*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

実績の調査：調査結果

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額(主に生産支援、研究開発目的を除く)の2013年からの累積額を示す。

年	米国	中国	韓国	台湾	ドイツ	欧州（ドイツ含）
2013	129.0	4,678.9	7.6		31.7	31.7
2014	303.0	13,684.2	13.6	0.8	76.3	76.3
2015	478.0	30,768.4	19.0	9.9	144.8	144.8
2016	548.0	67,535.2	23.8	407.3	205.1	263.8
2017	947.7	85,898.3	30.7	749.5	3,088.4	3,275.4
2018	1,177.7	90,224.8	65.6	1,544.9	3,131.9	3,318.8
2019	1,832.8	98,568.2	92.9	1,689.4	3,200.0	3,387.0
2020	1,972.8	110,062.3	466.7	1,873.8	3,247.0	3,433.9

単位は百万米ドル*1



(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

1-3-3. 調査結果分析

半導体生産能力、製造装置売上額ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸びと2020年末時点の能力、及び同じ期間での半導体製造装置売上金額を示す。

順位	公的支援金額 ^{*1}				生産能力 ^{*3}				半導体製造装置売上額 ^{*4}	
	研究開発含		生産能力拡充、その他 ^{*2}		伸び		2020年末時点の生産能力		国・地域名	金額 (百万米ドル)
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力 (Kウェーハ/月)		
1	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	台湾	97,780
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	韓国	88,930
3	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	72,610
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	北米	49,140
5	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	欧州	21,220
6	韓国	480.5	韓国	466.7						

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成

*4: SEAJ, SEMI, SEMIジャパンのデータを元にNTTアドバンステクノロジー株式会社が作成

半導体生産能力、製造装置売上額ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸びと2020年末時点の能力、及び同じ期間での半導体製造装置販売金額を示す。

順位	公的支援金額 ^{*1}				生産能力 ^{*3}				半導体製造装置売上額 ^{*4}	
	研究開発含		生産能力拡充、その他 ^{*2}		伸び		2020年末時点の生産能力		国・地域名	金額 (百万米ドル)
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力 (Kウェーハ/月)		
1 ①	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	台湾	97,780
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	韓国	88,930
3	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	72,610
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	北米	49,140
5	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	欧州	21,220
6	韓国	480.5	韓国	466.7						

- ① **【事実】中国の公的支援金額は他国・地域に比べると巨大で生産能力も最も伸びている。**
【考察】中国企業が巨大な公的支援を活用して生産能力を大きく伸ばしていると考えられる。

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成

*4: SEAJ, SEMI, SEMIジャパンのデータを元にNTTアドバンステクノロジー株式会社が作成

半導体生産能力、製造装置売上額ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸びと2020年末時点の能力、及び同じ期間での半導体製造装置販売金額を示す。

順位	公的支援金額*1				生産能力*3				半導体製造装置売上額*4	
	研究開発含		生産能力拡充、その他*2		伸び		2020年末時点の生産能力		国・地域名	金額(百万米ドル)
	国・地域名	金額(百万米ドル)	国・地域名	金額(百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力(Kウェーハ/月)		
1	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	台湾	97,780
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	韓国	88,930
3②	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	72,610
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	北米	49,140
5	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	欧州	21,220
6	韓国	480.5	韓国	466.7						

② 【事実】中国以外では、欧州や米国の公的支援金額が大きいですが、生産能力やその伸びは他国・地域よりも小さい。

【考察】米国や欧州は半導体の製造方法に関する研究開発の支援など、半導体そのものの売上に直結しない支援が大きいためではないかと考えられる。

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成

*4: SEAJ, SEMI, SEMIジャパンのデータを元にNTTアドバンステクノロジー株式会社が作成

半導体生産能力、製造装置売上額ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸びと2020年末時点の能力、及び同じ期間での半導体製造装置販売金額を示す。

順位	公的支援金額 ^{*1}				生産能力 ^{*3}				半導体製造装置売上額 ^{*4}	
	研究開発含		生産能力拡充、その他 ^{*2}		伸び		2020年末時点の生産能力		国・地域名	金額(百万米ドル)
	国・地域名	金額(百万米ドル)	国・地域名	金額(百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力(Kウェーハ/月)		
1	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	台湾	97,780
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	韓国	88,930
3	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	72,610
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	北米	49,140
5 ^③	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	欧州	21,220
6	韓国	480.5	韓国	466.7						

③ 【事実】韓国や台湾は公的支援金額は他国・地域より小さいが、半導体製造装置の販売金額や生産能力が大きい。

【考察】韓国企業や台湾企業は公的支援に頼らず企業努力で生産能力を高めてきたと考えられる。

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成

*4: SEAJ, SEMI, SEMIジャパンのデータを元にNTTアドバンステクノロジー株式会社が作成

半導体生産能力、製造装置売上額ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸びと2020年末時点の能力、及び同じ期間での半導体製造装置販売金額を示す。

順位	公的支援金額 ^{*1}				生産能力 ^{*3}				半導体製造装置売上額 ^{*4}	
	研究開発含		生産能力拡充、その他 ^{*2}		伸び		2020年末時点の生産能力		国・地域名	金額 (百万米ドル)
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力 (Kウェーハ/月)		
1 ①	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	台湾	97,780
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	韓国	88,930
3 ②	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	72,610
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	北米	49,140
5 ③	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	欧州	21,220
6	韓国	480.5	韓国	466.7						

- ① 中国の公的支援金額は他国・地域に比べると巨大で生産能力も最も伸びている。
 - ② 中国以外では、欧州や米国の公的支援金額が大きいですが、生産能力やその伸びは他国・地域よりも小さい。
 - ③ 韓国や台湾は公的支援金額は他国・地域より小さいが、半導体製造装置の販売金額や生産能力が大きい。
- ⇒韓国や台湾が公的支援に頼らず企業努力で生産能力を高めてきたのに対して、中国は巨大な補助金により生産能力を高めていることが推測される。

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成

*4: SEAJ, SEMI, SEMIジャパンのデータを元にNTTアドバンステクノロジー株式会社が作成

半導体生産能力、Fabless企業売上ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸び、及び2020年のFabless企業の売上金額を示す。

順位	公的支援金額* ¹				生産能力* ³				2020年の地域別Fabless企業売上* ⁴	
	研究開発含		生産能力拡充、その他* ²		伸び		2020年末時点の生産能力		国・地域名	金額 (十億ドル)
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力 (Kウェーハ/月)		
1	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	米国	81.9
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	台湾	23.0
3	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	19.2
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	欧州	
5	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	韓国	
6	韓国	480.5	韓国	466.7						

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成 *4: 出典: NTTアドバンステクノロジー株式会社

半導体生産能力、Fabless企業売上ランキングとの比較

- 情報収集方法で述べた方法に基づいてオープンソースから収集した、2013年-2020年の各国・地域の半導体に関する公的支援金額の総計額、同じ期間での生産能力の伸び、及び2020年のFabless企業の売上金額を示す。

順位	公的支援金額* ¹				生産能力* ³				2020年の地域別Fabless企業売上* ⁴	
	研究開発含		生産能力拡充、その他* ²		伸び		2020年末時点の生産能力		国・地域名	金額 (十億ドル)
	国・地域名	金額 (百万米ドル)	国・地域名	金額 (百万米ドル)	国・地域名	伸び(%)	国・地域名	生産能力 (Kウェーハ/月)		
1	中国	110,525.4	中国	110,062.3	中国	127.4	台湾	4,448	米国	81.9
2	欧州	15,419.8	欧州	3,433.9	韓国	57.5	韓国	4,253	台湾	23.0
3	米国	8,475.6	ドイツ	3,247.0	台湾	34.8	中国	3,184	中国	19.2
4	ドイツ	3,247.0	米国	1,972.8	北米	14.0	北米	2,623	欧州	
5	台湾	1,950.4	台湾	1,842.3	欧州	7.0	欧州	1,177	韓国	
6	韓国	480.5	韓国	466.7						

【事実】米国の生産能力は中韓台に比べて大きくは無いが、Fabless企業の売上は他国・地域に比べて大きく、2020年は全世界の2/3程度を占める。

【考察】米国企業は高度な半導体の技術開発などに公的支援を活用してFabless企業の売上を伸ばしてきた可能性がある。

(出典) NTTアドバンステクノロジー株式会社

*1: 2021年9月21日時点のレートで算出(1米ドル109円、1元17円、1ウォン0.09円、1台湾ドル3.9円、1ユーロ128円)

*2: 明確に研究開発支援目的であることが明記されている情報を除いた金額

*3: IC Insightsのデータを元にNTTアドバンステクノロジー株式会社が作成 *4: 出典: NTTアドバンステクノロジー株式会社

(参考) 主要企業の調査

企業候補

- 世界半導体売上上位企業の状況を整理する。各区分2企業程度を調査する。

企業名	区分	国・地域	順位*1	取得状況*2	OECD*3
Intel	IDM	米	1	△	○
Samsung Electronics	IDM	韓	2	X	○
TSMC	Foundry	台	3	○	○
SK Hynix	IDM	韓	4	○	○
Micron	IDM	米	5		○
Qualcomm	Fabless	米	6	X	○
Broadcom	Fabless	米	7	△	
NVIDIA	Fabless	米	8	X	○
Texas Instruments	IDM	米	9		○
Infineon Technologies	IDM	独	10	○	○
MediaTek	Fabless	台	11	X	
Apple	Fabless	米	13	X	
ST Microelectronics	IDM	欧	14		○
AMD	Fabless	米	15	△	

企業名	区分	国・地域	取得状況*2	OECD*3
UMC	Foundry	台	○	○
NXP	IDM	欧		○
Amkor	OSAT	米	X	○
ASE	OSAT	台	○	○
Hua Hong	Foundry	中		○
JCET	OSAT	中	○	○
SMIC	Foundry	中		○
Tsinghua (紫光集団)	Fabless	中	○	○
Vanguard Semiconductors	Foundry	台		○
HiSilicon	Fabless	中	■	

*1: 2020年売上ランキング
(<https://www.icinsights.com/news/bulletins/Intel-To-Keep-Its-Number-One-Semiconductor-Supplier-Ranking-In-2020>)
(12位は日本のKioxia)

*2: 財務諸表から公的支援金額が得られる場合○、一部の年のみの場合△、未記載X、財務諸表なし■、空白は未確認

*3: OECD調査対象企業
(<https://www.oecd.org/tokyo/newsroom/documents/Measuring-distortions-in-international-markets-ja.pdf>)

AMDの記載は2013と2014の本文のみ

調査結果

- 以下の企業について、財務諸表を基に生産支援に関連すると思われる補助金に関するデータを調査した結果を示す。

企業	Intel	SK Hynix	Infineon	TSMC	UMC	ASE	JCET	紫光集団
国・地域	米	韓	独	台	台	台	中	中
区分	IDM	IDM	IDM	Foundry	Foundry	OSAT	OSAT	Fabless
2013	129	6.3	31.7		0	5.4	9.3	1.5
2014	104	5.9	44.6	0	0.8	6.6	6.9	2.8
2015	105	5.5	47	0	9.1	6.3	9.3	1.1
2016		4.8	38.8	55.1	342.4	11.9	32.2	6.2
2017		4.8	43.4	100.5	241.9	12.2	54.7	124.6
2018	160	16.1	43.4	0	255.1	15.6	24.2	84.9
2019		15.3	68.1	122.4	22.1	22.3	46.2	123.1
2020		14.3	47	38.1	8.7	28.7	28.8	70.6
項目	Proceeds from government grants	Property, Plant, Equipment - Government grants - Land, Buildings, Structures, Machinery	Grants and subsidies - Cost of goods sold	Proceeds from government grants - property, plant and equipment, land use right and others	Government grants related to assets acquisition	Other income - Government subsidy	计入当期损益的政府补助	计入当期损益的政府补助

主要国・地域の税制

- 法人税や地方税を含めた実行税率は、米独が30%程度でほぼ同じ、中韓が25%、台湾が20%ともっとも小さい。
- 研究開発費に対しては各国・地域が控除を行っているが、中国の75%が目立つ。
- 半導体関連企業をターゲットにした優遇策を打ち出している国・地域があるが、特に中国は免税措置を打ち出しており、他国・地域の企業に比べ中国企業が有利になると思われる。

国・地域	米	中	韓	台	独
実効税率(%)	28	25	25	20	30
内訳(%)	連邦CIT: 21 地方CIT: 1-12	国CIT: 25	国CIT: 10-25	当局CIT: 20	法人税(国): 15.825 貿易税(地方): 8.75-20.3
優遇策など(半導体関連企業に影響するもの)	<ul style="list-style-type: none"> •最大20%程度の研究開発控除 •半導体製造分野への投資に対する25%の税額控除(2021年6月FABS法案提出) 	<ul style="list-style-type: none"> •75%の研究開発控除 •2020年、回路線幅によって数年間の免税や減税を発表(e.x. 28nm未満の場合10年間免税) 	<ul style="list-style-type: none"> •2%程度の研究開発控除(中小企業の場合25%) •1~20%の設備投資控除 •非メモリー分野のシステム半導体設計・製造技術研究費の約30%が控除対象(租税特例制限法) 	<ul style="list-style-type: none"> •最大15%の研究開発控除 •台湾で製造されていない機器や設備を輸入の際、関税免除 •スマート機器、5G関連項目について最大10億台湾ドル(39億円)控除 	<ul style="list-style-type: none"> •2019年までは優遇措置なし •2020年より特定の研究開発目的のための給与と賃金の25%が非課税、最大年間100万ユーロ(1.28億円)。

CIT: Corporate Income Tax(法人所得税)

(出典) NTTアドバンステクノロジー株式会社、<https://taxsummaries.pwc.com/>

【参考】中国の半導体関連企業に関する優遇税制

- 具体的な技術力に応じて、免税や半免措置を打ち出しており、国として積極的に先端技術の獲得を奨励していることが分かる。

条件	優遇税制
上海の深セン、朱海、汕頭、廈門、海南、浦東新区に設立された適格なハイテク企業	2年免税 + 3年半免
運用期間が15年を超える場合、線幅が 28 nm （両端を含む） 未満 の指定された集積回路製造企業またはプロジェクト	10年免税
運用期間が15年を超える場合に限り、線幅が 65nm （両端を含む） 未満 の集積回路を製造する指定された集積回路製造企業またはプロジェクト	5年免税 + 5年半免
運用期間が10年を超える場合に限り、線幅が 130nm （両端を含む） 未満 の集積回路を製造する指定された集積回路製造企業またはプロジェクト	2年免税 + 3年半免
奨励される集積回路設計/機器/材料/パッケージング/テスト企業	2年免税 + 3年半免
奨励され指定された鍵となる集積回路デザイン企業	5年免税

企業別税率について

- 以下の企業の2020年Annual Reportにおける税率を算出した。また、税率や税優遇に関する記載を調査した。
- 企業により記載内容に差異があるが、中台の企業は優遇策に関する記載があり、子会社を含めて積極的に活用していることが伺われる。

企業	Intel	SK Hynix	Infineon	TSMC	UMC	ASE	JCET	紫光集団
国・地域	米	韓	独	台	台	台	中	中
区分	IDM	IDM	IDM	Foundry	Foundry	OSAT	OSAT	Fabless
2020年税率 *1	16.7	23.7	12.3	12.6	7.5	18.1	8.8*2	14.5*3
2020年 Annual Reportなどの記載	実効税率は、中国NANDメモリ事業の売却などの影響で2020年に上昇し16.7%。	税率や優遇策に関する記載は見当たらなかった。	15%の法人所得税率に加えて、5.5%の連帯追加料金と12%の貿易税率に基づき法定税率28%。2020年は利益低下に伴い、実効税率12.3%。	2018年から、税法改正により法人所得税率が17%から20%に引き上げられ、未収金に課せられる付加税率が10%から5%に引き下げられた。研究開発費の10-15%の控除。 半導体生産施設の建設および生産能力の拡張から生じる所得、購入を目的とする増資に対して5年間免税。 代替ミニマム税（「AMT法」）のため12%。 生産施設の増資による免税は2020年約17億ドル。	2020年の実効税率は7.5%。	子会社(ASE Test)の拡張について免税。中国子会社が、研究開発費控除(15%)。	税制上の優遇措置が約2千萬元。 ハイテク企業認定により、15%の優遇税率。 韓国の子会社について、13-19%の優遇税率。	基本は25%。 ハイテク企業認定を受けた子会社は優遇税率15%。 香港の子会社は8.25-16.5%。

(出典) NTTアドバンステクノロジー株式会社

*1: 米韓台独は売上税(income tax)/税引前利益、中国は(調整後)所得税費用/純利益

*2: 調整前の税率は15%

*3: 調整前の税率は25%

1-3-4. 公的支援に関する最近の話題

1-3-4-1. 地域毎の公的支援に関する最近の話題

関連情報：公的支援に関する最近の話題

- 2020年以降、各国・地域とも数百億円～数兆円*1に上る支援を計画している。

	～2020	2021.01	2021.04	2021.07	2021.10	2022.01
米国		▲ 1月、NDAA2021成立 半導体生産設備建設1件当り最大30億ドル(3,300億円)の補助金		▲ 6月、United States Innovation and Competition Actが上院で可決、半導体生産工場の建設などに5年間で約520億ドル(5.7兆円)の補助金		▲ 2月、America COMPETES Act of 2022が下院で可決、金額は左記と同様。
中国	BigFund第2期(2019～)、総額2千億元(約3.4兆円)規模					
		▲ 8月、半導体産業に対する税制上の優遇策を発表(10年間法人税免除など)	▲ 3月、チップメーカーの輸入非課税を発表			
韓国		5月、K-半導体戦略 10年間で、K-半導体ベルトを構築、国内半導体業界への税制支援、R&D投資など(3年間で550億ドル～650億ドル(6～7兆円)との試算あり)	▲		▲ 11月、産業通商資源部は、半導体を含む産業の研究開発などに2兆6000億ウォン(2500億円)の資金投入発表	▲ 高度な技術者への補助金制度の創設を発表(特別手当の3割を政府が負担)
台湾	▲ 6月、チップメーカー誘致に100億台湾ドル(390億円)の補助金プログラム発表					
		▲ 9月、8世代半導体計画予算計上 海外メカの誘致や材料を台湾内で生産するために5年間で56億台湾ドル(218億円)				
ドイツ 欧州		▲ 12月、EU加盟国が、半導体の設計・製造に注力すること等を記した共同宣言に署名 2025年までに1450億ユーロ(19.1兆円)を目途に投資	▲ 4月、EUが半導体メーカーの欧州工場建設誘致に向けて、計100億ユーロ(1兆3,100億円)の補助金を出すとの報道		▲ 11月、EUは、加盟国政府が半導体事業に資金支援できるようにすることを検討すると表明	
				▲ 9月、独政府、半導体前工程工場誘致に50億～100億ユーロ(0.64～1.28兆円)の助成金との報道		

関連情報：米国の公的支援に関する最近の話題

- 米国では2020年7月のCHIPS法成立後、2021年も大規模な政府支援の動きが見られる。
- 2021年1月、CHIPS法(CHIPS for America Act(2020年7月))の趣旨を反映した国防権限法(National Defense Authorization Act, NDAA2021)が成立。半導体生産設備建設1件当たり最大30億ドル(約3,300億円)の補助金拠出を盛り込む。^{*1}
- 実際、サムスン電子は、CHIPS法により、固定資産税の減免を含めて、10億ドル以上のインセンティブを取得することができるとの予測も有る。^{*2} なお、サムスン電子に対しては、テキサス州やニューヨーク州が工場誘致に税制優遇など支援策を打ち出している。^{*2}
- 2021年6月、今後5年間で約520億ドル(約5.7兆円)を米国内において半導体生産工場の建設を行う企業等への補助金として拠出する米国技術革新・競争法案(United States Innovation and Competition Act) が上院で可決。^{*3} 2022年2月には下院で同様の内容の法案("America COMPETES Act of 2022")が可決された。^{*4}

*1: <https://www.nedo.go.jp/content/100931733.pdf>

*2: <https://www.etoday.co.kr/news/view/2042120>

*3: https://dl.ndl.go.jp/view/download/digidepo_11723354_po_084902.pdf?contentNo=1

*4: https://www3.nhk.or.jp/nhkworld/en/news/20220205_07/

関連情報：中国の公的支援に関する最近の話題(1/2)

- 中国では、3.4兆円規模のIC産業投資ファンド第二期が進行中など大規模な政府支援が行われている。
- 2020年8月、中国国務院は、28nm以下かつ15年以上運営されている半導体生産企業は10年間法人税免除などの半導体産業に対する税制上の優遇策を発表した。^{*1}
- 2021年3月、中国財務省などは、米国による半導体産業への制裁を受けて、チップメーカーが2030年まで非課税で機械や原材料を輸入できると発表した。ただし、具体的な補助金の額については発表されていない。^{*2}
- 2022年1月、中国政府が海外半導体大手も巻き込んだ共同開発拠点の誘致、地方政府と連携して買収資金を含めた資金提供を計画しているとの報道があった。^{*3}
- (参考) 2019年9月、HSMCという中国のFoundryが政府補助金を他のことに使うという詐欺が行われたとの報道^{*4}もあり、公的支援が適切に利用されていない事例も見られる。

*1: <https://www.china-briefing.com/news/china-integrated-circuit-software-enterprises-tax-incentives/>

*2: <https://www.businesstoday.in/current/world/china-announces-tax-breaks-to-spur-semiconductor-development/story/435162.html>

*3: <https://www.nikkei.com/article/DGXZQOGM283XQ0Y1A221C2000000/>

*4: <https://www.sedaily.com/NewsView/1ZAKZO3RWL>

関連情報：中国の公的支援に関する最近の話題(2/2)

- 中国における公的支援に関しては、全容が解明されておらず不透明との指摘も見られる。
- 中国IC産業投資ファンド(ビッグファンド)第1期(1,387億元(約2.4兆円))に続き、第2期のファンドが2千億元(約3.4兆円)規模で始まっており、多額の支援が進行中である。ただし、中国の中央・地方政府が企業にどの程度の資金的支援を行っているのかに関する全容の解明は進んでいない。^{*1}
- 公的支援の中には市場に比べて期待される利益を下回る水準で政府が投資を行う形態があり、主に中国に集中しているが、このような支援はあらゆる政府補助形態の中で定量的に測定することが最も難しい。^{*2}
- 中国では、鉄鋼などの産業部門とは異なり、どの半導体企業が国営企業であるか、政府投資であるかが明らかではない、との指摘もある。^{*3}
- また、2017年にはトランプ政権が中国の半導体企業に対する不透明な補助金について、不公正貿易を調査している。^{*4}

*1: <https://www.jri.co.jp/MediaLibrary/file/report/jrIREview/pdf/11597.pdf>

*2: <https://www.oecd.org/tokyo/newsroom/documents/Measuring%20distortions%20in%20international%20markets-ja.pdf>

*3: [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=TAD/TC\(2019\)9/FINAL&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=TAD/TC(2019)9/FINAL&docLanguage=En)

*4: <https://www.rieti.go.jp/jp/publications/nts/21e064.html>

関連情報：韓国への公的支援に関する最近の話題(1/2)

- 韓国は、2021年に「K-半導体戦略」を発表し、今後大規模な政府支援が見込まれる。
- 2021年5月には、半導体産業の競争力強化のための包括的な政策である「K-半導体戦略」を発表。2030年までに韓国に半導体集積地域(K-半導体ベルト)を構築し、安定した供給体制を整えること、製造装置や素材分野は外国企業の誘致を拡大すること、研究開発や施設整備に対する税額控除を拡充すること等のため、10年間で国内半導体業界の累積投資額が約510兆ウォン（約46兆円）になると想定し、これらに対する支援を行う。^{*1}
- 具体的には、R&Dに最大50%、製造に16%の税額控除、長期ローン8億8,600万ドル(約970億円)、政府によるR&D投資13億ドル、規制緩和、インフラストラクチャのアップグレードが含まれる。SIAは、韓国のチップ企業に対するこれらの新しい減税は、今後3年間で550億ドルから650億ドル(約6から7兆円)近くのインセンティブになる可能性があるの見積もっている。^{*2}
- 2021年1月に忠清南道庁とダイキンがエッチングガス生産工場建設で、2021年3月に忠清北道・清州市とASMが金属工場設置で、それぞれMOUを締結しており、その中には外国投資促進法に基づく補助金も含まれているケースがある。^{*3,4}

*1: https://dl.ndl.go.jp/view/download/digidepo_11723354_po_084902.pdf?contentNo=1

*2: <https://www.semiconductors.org/global-governments-ramp-up-pace-of-chip-investments/>

*3: <https://news.mynavi.jp/article/20210126-1674121/>

*4: <https://asm-au.com/asm-signs-mou-with-south-korean-provincial-government-for-metals-plant-location/>

関連情報：韓国への公的支援に関する最近の話題(2/2)

- 韓国は、2021年に「K-半導体戦略」を発表し、今後大規模な政府支援が見込まれる。
- 2021年11月、産業通商資源部は、半導体を含む5大産業分野の「素材・部品・装備特化団地推進団」発足式を開き、研究開発、テストベッド構築、専門人材養成などに対して2兆6,000億ウォン(約2,500億円)の資金を投入する計画を発表した。^{*1} 半導体については、SK Hynixが進める京畿道龍仁市の大規模半導体製造クラスターの建設が指定され、SK Hynixの巨大前工程ファブ4棟のほか、製造装置・材料メーカー、半導体設計企業などエコシステムパートナー約50社を収容する大規模な工業団地が数年後に完成する見込みである。
- (参考) 2019年には、ハンガリー政府が韓国のサムスンSDIに対して供与する計画の1億800万ユーロ相当の国家補助について、EUの国家補助ルールに照らしてその妥当性に疑義があるとして、欧州委員会が調査を行っている。^{*2}
- 2022年2月には、韓国政府が韓国で働く半導体技術者のDBへの登録を義務付ける計画を発表。同時に、技術者をつなぎとめるため、高度な技術者に支払う特別手当の3割を政府が負担する補助金制度の創設も盛り込む。

*1: <https://news.mynavi.jp/article/20211102-2175855/>

*2: <https://www.jetro.go.jp/biznews/2019/10/6250e7ff3d7d282b.html>

*3: <https://www.nikkei.com/article/DGXZQOGM138MK0T10C22A1000000/>

関連情報：台湾の関係機関による支援に関する最近の話題

- 台湾は2020年に工場誘致などの関係機関による支援を発表した。他国・地域に比べると金額の規模は小さい。
- 2020年6月、台湾当局は、台湾への投資を増やすために世界のチップメーカーを誘致するために100億台湾ドル(約390億円)の補助金プログラムを導入し、半導体会社が台湾に新工場を建設する場合、当局がそれを負担すると発表した。^{*1}
- 2020年9月、2021年～2025年を対象に、「A世代半導体計画」の予算として、経済部で37億台湾ドル(約144億円)、科技部で19億台湾ドル(約74億円)を計上。^{*2} 台湾の半導体設備の価値を高める、海外のメーカーを台湾に誘致しキーとなる設備を設置する、半導体材料の自主生産、などを目標として掲げる。
- (参考)2021年7月、APECでTSMC創業者のMorris Chang氏が各国政府による補助金支援による工場誘致をけん制する発言をしたと伝えられている。^{*3}
- (参考)2021年9月、台湾立法院(台湾最高の立法機関)の蔡其昌副院長が、中部科学園区管理局の許茂新局長らとともにTSMCの中部科学園区にあるFab15を訪問し、近隣の台中高爾夫球場(台中ゴルフカントリークラブ)や軍用地、財政部国有財産署の公有地など併せて約100haの用地への2nmファブの建設を打診したとの報道があった。^{*4}

*1: <https://ee.ofweek.com/2020-08/ART-8130-2800-30453949.html>

*2: <https://www.jetro.go.jp/biz/areareports/2021/dbd0fa7223039355.html>

*3: <https://www.mapion.co.jp/news/column/cobs2270731-1-all/>

*4: <https://news.mynavi.jp/article/20210906-1965957/>

関連情報：ドイツ、欧州の公的支援に関する最近の話題(1/2)

- 欧州では2020年末から2021年にかけて大規模な政府支援を発表、報道されている。
- ドイツ政府としても大規模な助成金を計画している。
- 2020年12月、ドイツ、フランス、イタリア等17のEU加盟国が、半導体の設計・製造に注力すること等を記した共同宣言に署名した。2025年頃までに、欧州復興基金の20%、1,450億ユーロ(約19.1兆円)を目途に投資を行うとしている。^{*1}
- 2021年3月、2030年までの次世代デジタル産業の育成方針「デジタルコンパス」を発表した。この中で、半導体については生産の世界シェアを現在の2倍となる20%に引き上げるという目標が掲げられている。^{*1}
- 2021年4月、EUがTSMC、米インテル、韓国サムスンの3大メーカーに欧州での工場建設誘致に向けて、計100億ユーロ(約1兆3,100億円)の補助金を出すとの報道があった。^{*2} 一方、インテルは工場建設に約1兆円を要求したとも言われている。^{*3}

*1: https://dl.ndl.go.jp/view/download/digidepo_11723354_po_084902.pdf?contentNo=1

*2: <https://www.nna.jp/news/show/2182783>

*3: <https://news.mynavi.jp/article/20210506-1884338/>

関連情報：ドイツ、欧州の公的支援に関する最近の話題(2/2)

- 欧州では2020年末から2021年にかけて大規模な政府支援を発表、報道されている。
- ドイツ政府としても大規模な助成金を計画している。
- 2021年9月、ドイツ政府経済相は、ドイツ政府として半導体サプライチェーン再構築に向け、50億～100億ユーロ(半導体工場を建設するコストの20%から40%)の助成金を出し、複数の半導体前工程工場を誘致する計画であると述べた。^{*1}
- 2021年11月、EUは国家補助のルールを緩和し、加盟国政府が半導体事業に資金支援できるようにすることを検討すると表明した。^{*2} これにより、研究開発目的だけでなく、欧州においても生産能力向上目的の公的支援が増加し、数年後、欧州の生産能力が大きく向上する可能性がある。
- 2022年2月、欧州委員会は「European Chips Act (欧州半導体法案)」の草案を公表した。法案が正式に成立すると、政府と民間からの430億ユーロ(5兆5,040億円)を超える投資、及び現在進められている研究、開発、技術革新の強化にも110億ユーロ(1兆4,080億円)の投資が見込まれる。^{*3}

*1: <https://news.mynavi.jp/article/20210906-1965957/>

*2: <https://www.nikkei.com/article/DGXZQOGR18EO10Y1A111C2000000/>

*3: <https://jp.taiwantoday.tw/news.php?unit=148,149,150,151,152&post=214734>

1-3-5. まとめ

総括

- これまで、中国が他国・地域に比べて巨大な補助金を活用して、生産能力を伸ばしてきた。
- 最近になって、米韓欧が大規模な公的支援を決定すると共に、中国も免税措置などの支援を決め、半導体企業に対する公的支援の環境が整いつつある。
- 台湾は他国・地域に比べると関係機関による支援の規模が小さいが、これは主要企業がTSMCとUMC程度に限られるためではないかと考えられる。

	補助金	生産能力	税制	主な最近の話題
米	欧州と並んで中国に次ぐ規模	伸び・生産能力共に欧米に次いで小さい⇒公的支援は研究開発中心と推定される	2021年に 半導体製造分野の投資に25%の税額控除法案提出	2021年上院、2022年下院で 5.7兆円規模の補助金に関する競争法案が成立
中	他国・地域に比べると巨大 全容は解明されておらず不透明	最も伸びている⇒巨大な公的支援を活用したと推定される	2020年、 半導体関連企業の優遇免税発表 、研究開発費控除75%	2019年から 3.4兆円規模のIC産業投資ファンド で多額の支援
韓	最も小さい	生産能力は台湾に次ぐ大きさ⇒公的支援に頼らず企業努力によるものと推定される	20%の設備投資控除、 システム半導体R&Dの30%控除 、2021年、 R&Dに最大50%、製造に16%の税額控除発表	2021年、K-半導体戦略発表、 3年間で6-7兆円の減税規模と試算
台	韓国に次いで小さい	生産能力は最も大きい⇒関係機関による支援に頼らず企業努力によるものと推定される	一部機器・設備の関税免除、スマート機器、5G関連項目について最大39億円控除	2020年、メーカ誘致に合計600億円規模の補助プログラム、予算計上
欧(独)	米国と並んで中国に次ぐ規模	伸び・生産能力共に最も小さい⇒公的支援は研究開発中心と推定される	2020年より特定の研究開発目的の給与と賃金の25%非課税	2021年、EUや独政府が 半導体工場建設誘致に1.3兆円規模の補助金 との報道

展望

- 米中韓欧が2020年以降発表した様々な公的支援を今後実行に移すことで、各国・地域が優位になるための半導体企業に対する公的支援の競争が激化すると想定される。
- 台湾は、主要企業であるTSMCが他国・地域の公的支援を活用して他国・地域での工場建設を進めている。一方、台湾当局は最先端技術は台湾の外へ出さない方針との報道もあり*1、台湾当局による今後の関係機関による支援策が注目される。

	展望
米	成立した 競争法案が具体化 することで 大規模な公的支援が実行 される
中	巨大な補助金に加え、 半導体関連企業への免税策 により 中国企業が益々有利 になる
韓	これまでは小さかったが、 K-半導体戦略の実現 に向けて、 大規模な公的支援が実行 される
台	他国・地域に比べて関係機関による支援の規模は小さい。 TSMCが他国・地域の公的支援を受けて他国・地域に工場を誘致 される
欧 (独)	WTO協定を遵守した国家補助ルールにより補助金が制限されていたが、 半導体産業では制限が緩和され、大規模な公的支援が実行 される

なお、巨大な公的支援が**企業の競争力を損ねる**という懸念や、各国・地域の企業が**他国・地域の企業への補助金に異議を唱えたり**、**WTOが公正な競争環境の整備のためのルールを改める**、など巨大な公的支援の方向性が変わる可能性も考えられる。

2. マイクロエレクトロニクスに係る技術動向

2-1. 世界のマイクロニクスに係る開発・応用技術に係る先端的技術の開発動向の調査	249
2-1-1. 調査の目的と対象国際会議等	250
2-1-2. VLSI Symposia 2021	252
2-1-3. Hot Chips 33	306
2-1-4. Arm DevSummit 2021	349
2-1-5. NVIDIA GTC 2021 Fall	386
2-1-6. SC21	422
2-1-7. RISC-V Summit 2021	440
2-1-8. IEDM 2021	492
2-1-9. ISSCC 2022	527
2-1-10. まとめ	562

2-1.世界のマイクロニクスに係る開発・ 応用技術に係る先端的技術の開発 動向の調査

2-1-1. 調査の目的と対象国際会議

【目的】

- 世界のマイクロエレクトロニクスに係る開発・応用技術に係る先進的技術の開発動向を調査する。

【対象国際会議】

- マイクロエレクトロニクスに関連する下記国際会議を調査対象とした。

#	会議名	開催地	開催期間
2-1-2	VLSI Symposia 2021	オンライン開催	2021年6月13日 - 19日
2-1-3	Hot Chips 33	オンライン開催	2021年8月22日 - 24日
2-1-4	Arm DevSummit 2021	オンライン開催	2021年10月19日 - 21日
2-1-5	NVIDIA GTC 2021 Fall	オンライン開催	2021年11月8日 - 11日
2-1-6	SC21	St. Louisとオンラインのハイブリッド開催	2021年11月14日 - 19日
2-1-7	RISC-V Summit 2021	サンフランシスコとオンラインのハイブリッド開催	2021年12月6日 - 8日
2-1-8	IEDM 2021	サンフランシスコとオンデマンドのハイブリッド開催	2021年12月11日 - 15日
2-1-9	ISSCC 2022	オンライン開催	2022年2月19日 - 26日

2-1-2. VLSI Symposia 2021

目次

1. 開催概要
2. VLSI シンポジウム2021概要
 - 2-1. VLSI Technologyシンポジウム2021概要
 - 2-2. VLSI Circuitsシンポジウム2021概要
3. プログラム
 - 3-1. 基調講演
 - 3-2. フォーカスセッション
 - 3-3. パネルディスカッション
 - 3-4. ショートコース
 - 3-5. フォーラムおよびワークショップ
4. Technologyシンポジウム2021および注目論文の全体概要
 - 4-1. Technology 特筆すべき論文
 - 4-1-1. フォークシート構造のN/Pトランジスタ技術
 - 4-1-2. 2D材料チャネルMOSFET技術
 - 4-1-3. 最先端3D NAND Flashメモリ
 - 4-2. Technology プレス発表されたその他注目論文
5. Circuitsシンポジウム2021および注目論文の全体概要
 - 5-1. Circuits 特筆すべき論文
 - 5-1-1. サーバ電源向けGaNパワーモジュール
 - 5-1-2. 短距離・高速チップ間有線通信
 - 5-1-3. 学習可能なエッジAIチップ
 - 5-2. Circuits プレス発表されたその他注目論文

1. 開催概要

1. 開催日時

2021年6月13日 - 19日 (7日間)

※オンデマンドコンテンツは8月31日まで利用可能

2. 開催形式

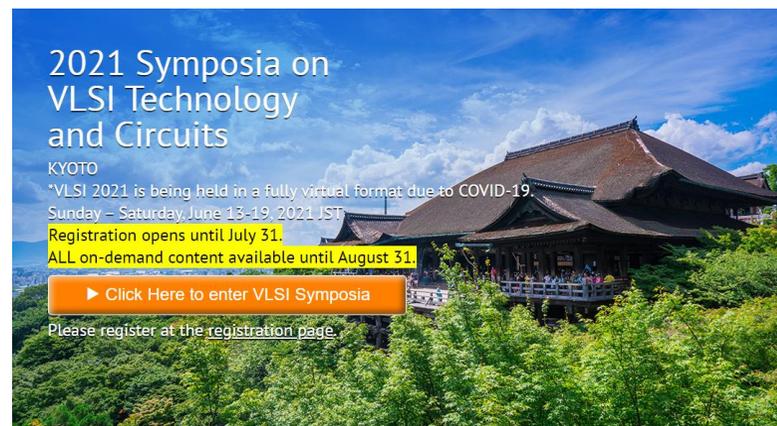
Virtual Conference (オンライン学会) で開催
バーチャルであっても参加者および講演者間の議論を深めるために
「オンデマンドプログラム」と「ライブプログラム」を用意

3. 構成

Technologyシンポジウム、Circuitsシンポジウムから構成され、各々およびジョイントで

- ・基調講演
- ・技術セッション
- ・フォーカスセッション
- ・パネルディスカッション
- ・ショートコース
- ・フォーラム
- ・ワークショップ

が記載された。



URL <https://vlsisymposium.org/>

2. VLSI シンポジウム2021概要

VLSIは毎年TechnologyとCircuitsの共通テーマを掲げており、VLSIシンポジウム2021では“*VLSI Systems for Lifestyle Transformation*”「**ライフスタイル変革のためのVLSIシステム**」と題して、世界的なCOVID-19のパンデミックに対する懸念が続き、新しいライフスタイルへの変革が期待される中、人工知能や機械学習、バイオメディカルやIoTの進化に繋がる最先端のプロセステクノロジーからシステムオンチップまで、様々なデバイス・画素・システムの新技术を取り上げ、それらの相乗効果の発揮を議論する機会を提供している。

今年の日本と中国の投稿数と採択率を比較すると、Technology、Circuitsともに投稿数では中国が多いが、採択率では日本が高くなっている。また、日本の採択率は全体での採択率よりも高く、中国は低い。現時点では日本のマイクロエレクトロニクス分野での技術的な存在感は中国よりも高いと言える。

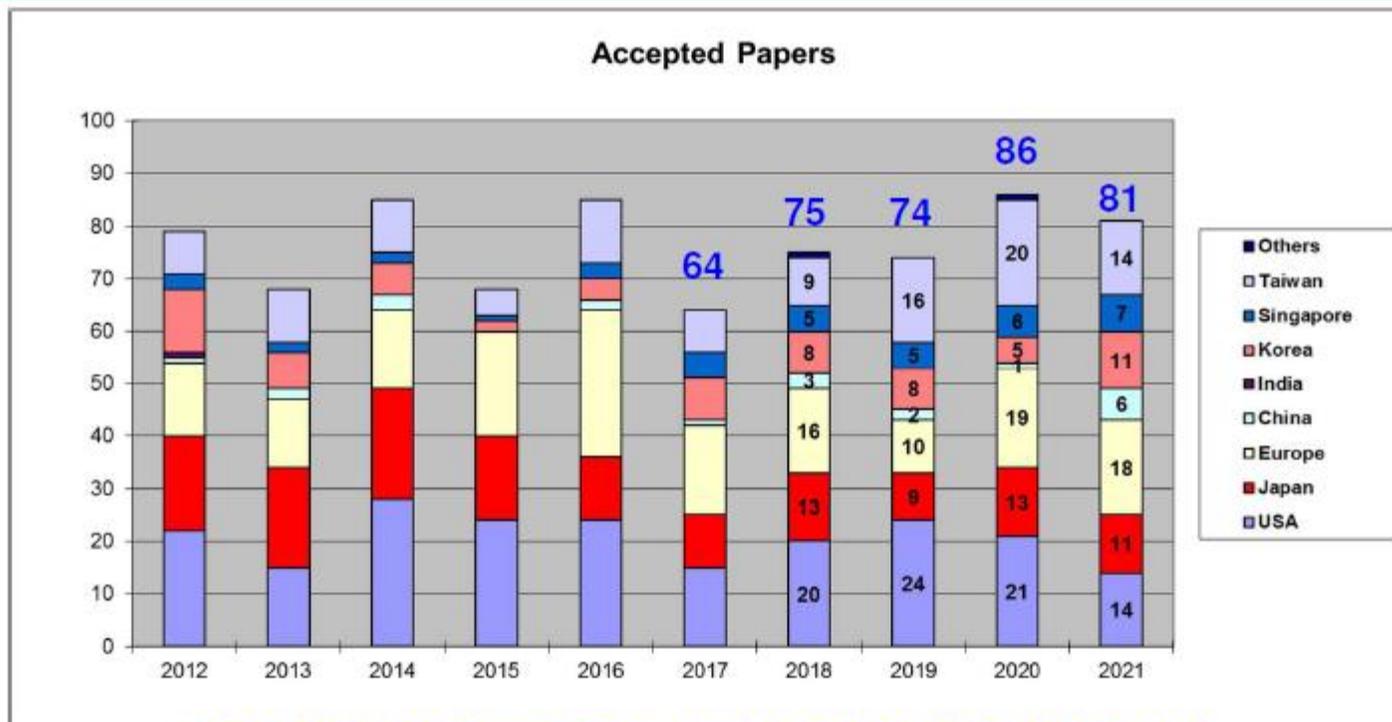
※投稿数・採択数の年次推移を次スライド以降に示す。

シンポジウム	区分	投稿数	採択数	採択率
Technology	全体	189	81	42.9%
	日本	19	11	57.9%
	中国	26	6	23.1%
Circuits	全体	300	102	34%
	日本	12	5	41.7%
	中国	31	6	19.4%

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

2-1. VLSI Technologyシンポジウム2021概要 (1)

地域別採択論文件数推移



採択論文件数で見ると、日本は地域別・国別4位。

1位: 欧州(18件※内ベルギー14,フランス3)、2位: 米国と台湾(14件)、
4位: 日本と韓国(11件)、6位: シンガポール(7件)、7位: 中国(6件)

2-1. VLSI Technologyシンポジウム2021概要 (2)

機関別採択論文数推移

Late Newsは除く

	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012
	Acc									
Total	81	86	74	75	64	85	68	85	68	79
imec	14	9	7	10	7	10	9	10	9	8
Samsung	7	5	4	6	6	1	0	1	4	6
The Univ. of Tokyo	7	3	3	3	3	2	4	4	4	4
TSMC	6	9	4	0	3	4	2	2	2	2
National Univ. of Singapore	6	4	4	4	3	1	0	1	2	3
Institute of Microelectronics, Chinese Academy of Science (IMECAS)	5	1	1							
IBM	3	5	8	4	6	7	6	5	3	7
Pohang University of Science and Technology (POSTECH)	3	0	1							
CEA-LETI	2	5	0	3	4	5	3	3	1	3
Macronix	2	3	3	2	0	2	3	4	5	4
National Taiwan Univ.	2	3	2	1	0	0	0	3	0	0
GLOBALFOUNDRIES	2	3	1	5	5	3	0	3	0	0
Univ. Notre Dame	2	1	4	2	2	1	0	0	0	0
Sony Semiconductor Solutions	2	1	0	0	1	0	0	1	0	0

昨年との比較: imec(ベルギー)が5増、東京大学とIMECAS(中国)とが4増、POSTECH(韓国)が3増。

2-1. VLSI Technologyシンポジウム2021概要 (3)

各国における採択論文種別

	Adv. CMOS	Beyond CMOS	Phys. Char.	Non-Si	Analog RF	Process	Reliab.	Memory	DTCO	3D	IoT	New Compute	Others
日本	③	2		1		③	①	5				②	3
韓国	1	1		1				5				2	1
台湾	1					1	1	8		1		2	
シンガポール	1	2		1		1		2					
中国			2	1			1	2					
米国	2	1	2	2	1	1	1	1	1			2	
ベルギー	2	2		1			1	3	2	1		2	
フランス	1							1		1			
ドイツ								1					

- ① 世界的に見て、昨年よりも、更に多くのメモリ関連論文が発表される。(20件→28件)
- ② New Computing関連論文が多いことは今年の特徴 (11件)。
- ③ 先端CMOSデバイス・プロセスに関する論文は、今年は減少傾向。(22→11件)

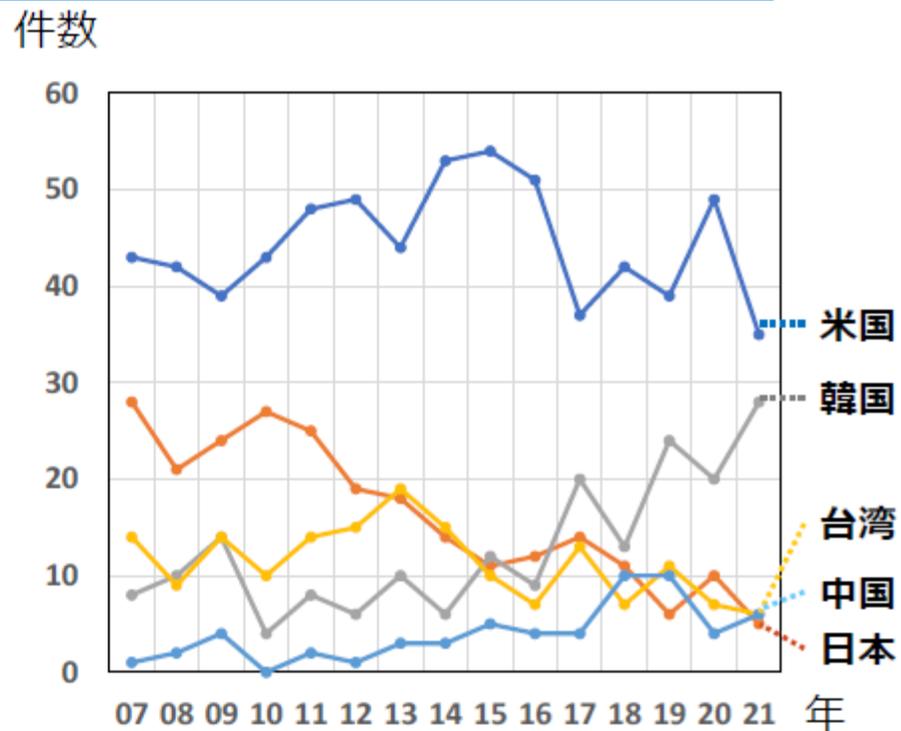
2-2. VLSI Circuitsシンポジウム2021概要（1）

採択論文の国別構成 ～牽引国～

首位米国の採択件数が34%。韓国が引き続き2位で27%の採択率を占める。日本は5件の採択となり一昨年と同様の5位に留まる。

国名	採択数	割合
米国	35	34%
韓国	28	27%
台湾	6	6%
中国	6	6%
日本	5	5%
シンガポール、ベルギー	各4	各4%
スイス、アイルランド	各3	各3%
その他	9	
Total	102	

※招待論文4件を除く



2-2. VLSI Circuitsシンポジウム2021概要 (2)

採択論文の機関別構成 ~牽引機関~

韓国KAISTが採択件数を大きく伸ばして首位。続いてSamsung、Intelが牽引する。
国内からは東京大学 2件、東京工業大学 2件 など。

2017		2018		2019		2020		2021	
ミシガン大	13	ミシガン大	8	ミシガン大	8	Intel	12	KAIST	14
KAIST	11	Intel	7	Intel	7	KAIST	9	Samsung	8
TSMC	5	カリフォルニア大学	8	KAIST	7	ミシガン大	7	Intel	7
国立台湾大	4	KAIST	6	UCバークレー	5	Samsung	5	ミシガン大	7
テキサス大 (オースチン)	4	ザイリンクス	5	Samsung	4	UCサンディエゴ	4	UC ロサンゼルス	4
デルフト工科大	3	IBM	4	ソウル大	4	バージニア大	4	清華大学	4
IBM	3	TSMC	3	TSMC	4	imec	4	imec	3
IMEC	3	清華大学	3	延世大	4	IBM	3	アイルランド大 ダブリン	3
国立清華大	3	中国科技大	3	香港科技大	3	TSMC	3	Media Tek	2
ルネサス	3	MIT	3	国立交通大	3	ルーヴェン・カトリック大	3	TSMC	2
				ルネサス	3			東京大学	2
				テキサス大 (オースチン)	3			東京工業大学	2
				テキサス大 (ダラス)	3			その他	1~2
全体	115	全体	104	全体	108	全体	110	全体	102

2-2. VLSI Circuitsシンポジウム2021概要 (3)

分野別採択論文数と日本の採択数

年		2017		2018		2019		2020		2021	
		全体	日本	全体	日本	全体	日本	全体	日本	全体	うち日本
デジタル	プロセッサ・アーキテクチャ	16	2	15	0	16	2	18	3	18	1 (PFN) *招待論文
	デジタル回路	6	2	8	0	6	0	11	0	6	1 (早稲田)
	クロック・周波数生成	11	2	9	1	8	1	10	1	7	1 (東大)
通信	無線通信	6	2	7	2	11	1	8	2	7	2 (東工大x2)
	有線通信	13	1	13	1	14	0	7	0	9	
メモリ	SRAM & DRAM	7	2	8	1	8	2	14	3	10	
	不揮発メモリ・ROM										
アナログ	データコンバータ	17	0	11	2	11	0	10	0	7	
	パワーマネジメント	7	0	10	1	10	0	8	1	11	1 (東大)
	フィルタ、アンプ	5	0	4	0	3	0	6	0	3	
	センサ、ディスプレイ	27	3	19	2	21	0	12	2*	12	1 (ルネサス) *招待論文
	バイオメディカル							6		11	
合計		115	12	104	10	108	6	110	6	102	5 + 招待論文 2

*招待論文

3. プログラム

3-1. 基調講演

TechnologyとCircuitsでそれぞれ2名が基調講演を行った。
講演タイトルと講演者を以下に示す。

■ Technology

講演タイトル	講演者/所属
“Pandemic Challenges, Technology Answers” 「パンデミックの課題とTechnologyの回答」	Dr. Siyoung Choi/President, GM of Foundry Business, Device Solutions Division, Samsung
“Materials to Systems in Semiconductor Manufacturing and Beyond” 「半導体製造における材料からシステムまで」	Dr. Om Nalamas/SVP and CTO, Applied Materials, Inc. President, Applied Ventrues, LLC

■ Circuits

講演タイトル	講演者/所属
“Fugaku and A64FX: the Fastest and the most Power Efficient General Purpose Arm CPU for the Fastest Supercomputer in the World” 「富岳とA64FX: 世界最速かつ最もパワー効率の高いArm CPU」	Prof. Satoshi Matsuoka/Director, Riken Center for Computer Science
“A New Era of Tailored Computing” 「新時代における特注のコンピューティング」	Mr. Mark Papermaster/CTO and EVP, Technology and Engineering, AMD

3-2. フォーカスセッション

Technology/Circuitsジョイントで5セッション、Technologyで2セッション、Circuitsで1セッション行われた。セッションタイトルを以下に示す。セッションは、一般投稿論文および招待講演論文(ある場合)から構成されている。

■ ジョイント

セッションタイトル	招待講演者(所属)
3D/Heterogeneous Integration (3D/異種 技術集積)	Geet Van der Plas(imec), Takafumi Fukushima(東北大)
Photonics Interconnect and Compute (Siフォトニクス)	Mark Wade(Ayar Lab), Nobuhiko Nishiyama(東工大)
Circuits and Technology for Quantum Computing (量子コンピューティング)	Iuliana Radu(imec), Yutaka Tabuchi(理研)
Computing-in-Memory (メモリーコンピューティング)	-
Image Sensors (イメージセンサー)	-

■ Technology

セッションタイトル	招待講演者(所属)
Advanced Memory Technology	Kay Yakushiji/Yuasa Shinji(産総研), Stefan Müller(Ferroelectric Memory GmbH)
New Process and Material for Future Devices	Chih-Pin Lin/Tuo-Hung Hou(国立陽明交通大学), Kevin Moraes(AMAT)

■ Circuits

セッションタイトル	招待講演者(所属)
Energy-Efficient Machine Learning Processors	K. Namura(Preferred Networks)

3-3. パネルディスカッション

Technology/CircuitsジョイントおよびTechnology、Circuitsそれぞれで開催された。

■ ジョイント

タイトル	The New Normal … How will it change work, life and education? (ニューノーマル: それは 仕事と生活と教育をどうかえるのか?)
司会進行	Dr. Kazuo Yano, 日立
内容	ポストパンデミック時代における社会環境の変化を予測する。VLSI業界以外の方からパネリストを招き、ニューノーマルについて語っていただく。

■ Technology

タイトル	3D/Heterogeneous integration: Are we running towards a Thermal Crisis? (3D/異種 集積技術: 熱的危機に対してどう対処していくのか?)
司会進行	Prof. Takayuki Ohba, 東京工業大学
内容	3D集積や異種集積をする上で、熱的問題がいつも話題にはなるが、学会等でもここに真正面から切り込んだ論文は少ない。半導体と熱の専門家の意見を聞く。

■ Circuits

タイトル	New Generation Chip Makers vs. the Incumbents (新世代のチップメーカー 対 伝統的なチップメーカー)
司会進行	Prof. Naveen Verma, Princeton University
内容	従来、システムメーカーは顧客として、VLSIチップメーカーからチップを購入してきたが、その構造に最近変化が見られ、自力でチップを作成する会社が出てきた。両者に登壇していただく。

3-4. ショートコース

Technology/Circuitsジョイント、Technology、Circuitsそれぞれで開催された。
タイトルを以下に示す。

主催	タイトル
ジョイント	Enabling a Future of Even More Powerful Computing (もっとパワフルなコンピューティング技術の将来を可能にする)
Technology	Advanced Process and device technology toward 2nm-COMS and emerging memory (2nm CMOSと新メモリにおける最先端プロセス・デバイス技術)
Circuits	Advanced Circuits and Systems for Internet-of-Things(IoT) Sensors (IoTセンサーのための最先端回路技術と最先端システム技術)

3-5. フォーラムおよびワークショップ°

フォーラムのタイトル・テーマおよび、公募によって決まった4つのワークショップテーマを以下に示す。

■ フォーラム

タイトル	テーマ
“Technologies for POST Covid-19 era” 「ポストCOVID-19時代のテクノロジー」	(1) Medical/Healthcare (医療/健康)
	(2) Communication (通信)
	(3) Security (安全)
	(4) Smart manufacturing/Logistics

■ ワークショップ°

	テーマ
ワークショップ1	AI/Machine Learning for Circuit Design and Optimization (回路設計と回路最適化のためのAI/機械学習)
ワークショップ2	PPAC Analysis and System-Technology Co-optimization for 3D Memory-on-Logic IC, Many-core SOC and AI Computing Applications (3次元メモリオンロジックIC、多コアSOC、AI経産用途のためのPPAC分析とシステム-テクノロジー協調)
ワークショップ3	Deep Analysis Can Compress the Time to Design Optimum Analog/Mixed-Signal Circuits (深層解析により、最適なアナログ/ミックスドシグナル回路の設計時間が短縮される)
ワークショップ4	Materials Introductions – A path forward for all devices (新材料導入: 全ての新デバイスのために)

4. Technologyシンポジウム2021および注目論文の全体概要

● Technology領域での世界の流れ

この領域は、CMOS LogicとMemoryに大きく二分される。現在製造されている最新Logicに使用されているFinFETの次として、NanoWire, NanoSheet, Forksheet等と呼ばれるGate All Around FETの研究が進められている。Memoryでは3D NANDの微細化が推し進められると共に、各種の新規Memoryの研究が、大学・研究機関含めて幅広くなされている。なお、LogicとMemory以外では、Sensor、Photonics、3D Package、New Computing向けDeviceがTechnologyのその他の注目領域である。

● VLSIシンポジウムでの発表の動向

最先端・次世代のLogicと3D NANDが本学会で報告されていることは、まず第一の特徴である。その他、特筆すべきは、投稿・採択が多かった分野が強誘電体デバイス・メモリ(FeRAM)であり、3つもセッションを組んでいる点である。未だ実用化されていない低電圧のFeRAMが、次世代メモリとして期待されていることの証である。また、New Computingや量子Computingのためのデバイスが報告されている点も今回の特徴である。

● 将来的な研究開発の方向性

全く異なる材料・素子を複数集積するHeterogeneous Integration(異種集積化)は高機能化を可能にし、今後のトレンドになる。また、3次元実装(3D Packaging)は素子の高性能化の鍵となる。さらに、New Computing向けDeviceのような特殊な用途を意識したDevice設計も研究開発が加速すると思われる。

● 注目論文全体概要

LogicもMemoryも、高性能化・微細化を推し進める論文が注目を集める。広義のMooreの法則は未だ終焉を迎える気配に無い。AIやHPCや5GといったApplicationが、次世代の研究開発を後押ししていると言える。

4-1. Technology 特筆すべき論文

インパクトのある論文としてVLSI Technologyシンポジウムの注目論文としての11件が紹介されており、その中から2件（T2-1, T2-3）、11件以外から1件（T2-2）の下表の3件を特筆すべき論文としてピックアップした。それぞれの詳細について解説する。

※プレス発表されたその他注目論文の9件については、リストを掲載する。

論文番号	論文タイトル	著者所属機関
T2-1	Forksheet FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space	imec
T2-3	Advancing Monolayer 2D NMOS and PMOS Transistor Integration From Growth to van der Waals Interface Engineering for Ultimate CMOS Scaling	Intel
T2-2	Highly Manufacturable 7 th Generation 3D NAND Flash Memory with COP structure and Double Stack Process	Samsung

4-1-1. フォークシート構造のN/Pトランジスタ技術 (1/4)

T2-1

Forksheets FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space

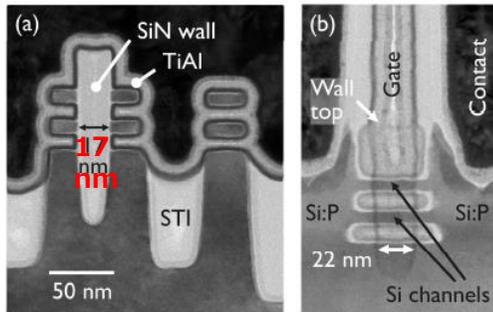
H. Mertens, R. Ritzenthaler, Y. Oniki, B. Briggs, B.T. Chan, A. Hikavy, T. Hopf, G. Mannaert, Z. Tao, F. Sebaai, A. Peter, K. Vandersmissen, E. Dupuy, E. Rosseel, D. Batuk, J. Geypen, G. T. Martinez, D. Abigail, E. Grieten, K. Dehave, J. Mitard, S. Subramanian, L.-Å. Ragnarsson, P. Weckx, D. Jang, B. Chehab, G. Hellings, J. Ryckaert, E. Dentoni Litta, N. Horiguchi
imec, Leuven, Belgium, email: hans.mertens@imec.be

(1) 本論文が応用される分野、製品

2nm世代以降のCMOS技術に適用予定。2nm世代は、5G以降のMobile/高速通信、データセンター等のHPC、AI/Deep Learningに用いられる。

(2) 当該分野における技術ニーズ

超高速・低消費電力・高集積度のMOSFETが、2nm世代CMOSには必要とされている。
超高速・低消費電力の為に導入した最先端のNanosheetと組み合わせ、高集積度化のための(本発表)Forksheets構造が必要となる。



“Forksheets FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space”, H. Mertens et al., imec

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

4-1-1. フォークシート構造のN/Pトランジスタ技術 (2/4)

(3)本論文の概要

16nm世代から導入されたFinFETは14/10/7/5nm世代と使われてきたが、そろそろ性能に限界を迎える。4nmあるいは3nm世代からは、更なる超高速・低消費電力を実現するNanosheetと呼ばれるGate All-Around(GAA)構造が用いられる予定である。

更に、2nm以降もNanosheetを継続使用するが、高集積度のためには、(昨年発表の)Buried Power Rail(BPR)と本発表のForksheetsheet構造が必要となってくる。

本論文では、Forksheetsheet構造において、ゲート長22nmまで短チャネル効果を抑制(SS=66~68mV/dec)、チャネル形成後のウェットクリーニングを最適化し、トランジスタのオンオフ特性を改善した。

また、従来構造によるスケーリングの難点であるDual work function metal gatesの間隔を**17nmまで縮小**した。

(4)技術的に特筆すべき点

以下の(a)-(c)の3点が特筆すべき点である。

(a)将来的に5-track Standard cellを実現する上で、Nanosheetでは、N-P 距離が足枷になって、NFET幅とPFET幅を十分に広く設定できない為、結果的に性能が向上しない。一方、Forksheetsheet構造では、N-P境界に壁を設定する事で、**N-P距離を極限まで短くできる**ことにより、NFET幅とPFET幅を十分に拡大することが可能となり、素子性能が向上する(Fig. 1)。

Nanosheetでは、N-P距離を近づけると、Maskアスペクト比とWet etch mask undercutが問題となり加工できないが (Fig2上段)、Forksheetsheet構造では、N-P距離を近づけた場合でも、上記の二つの問題が解決される (Fig2 下段)。

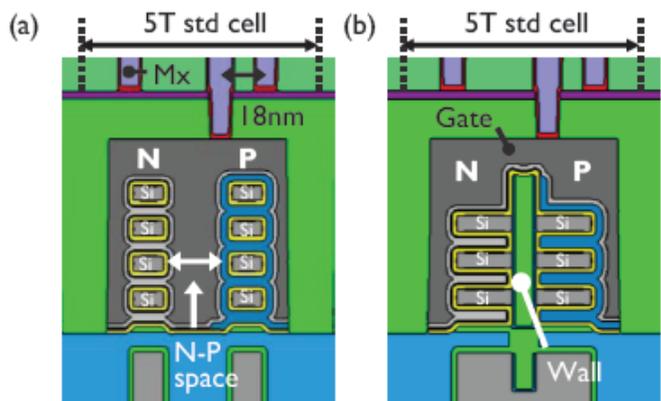


Fig. 1. (a) Nanosheet- and (b) forksheet-based 5-track standard cell

©JSAP

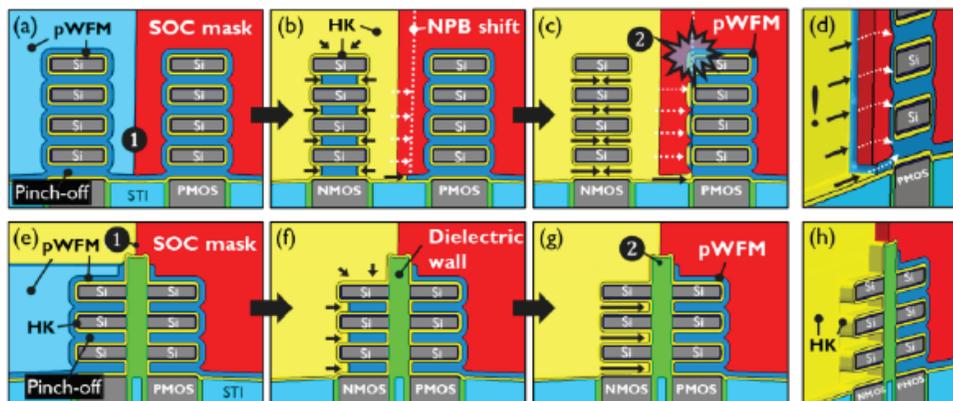


Fig. 2. Work function metal (WFM) Patterning for (a-d) Nanosheet and (e-f) forksheet. ©JSAP

4-1-1. フォークシート構造のN/Pトランジスタ技術 (3/4)

(b) N-P距離17nmのForksheets構造の試作を**実現**。N-MetalのTiAlを有したNFETとP-MetalのTiNを有したPFETの双方で、良好な I_d - V_g 特性が得られた。 $L_g=22\text{nm}$ において、 $SS=66\sim 68\text{mV/dec}$ である。

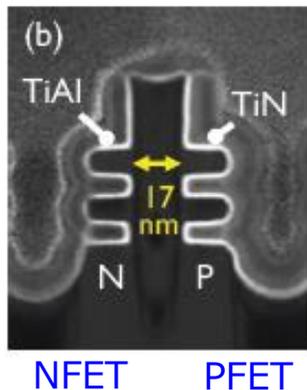


Fig. 6. X-section of Forksheet with TiAl (n-Metal) and TiN (p-Metal)
©JSAP

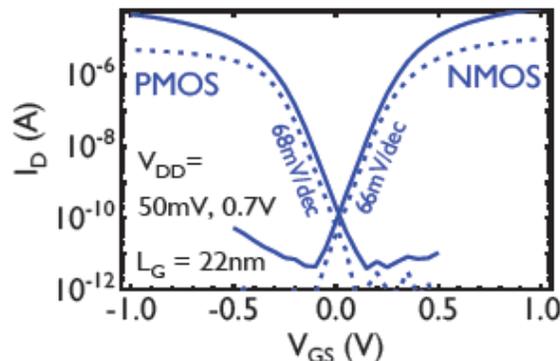


Fig. 7. I_d - V_g characteristic of Forksheet nFET and nFET with $L_g = 22\text{nm}$ ©JSAP

(c)従来からのNanosheetと、本論文のForksheetsを隣接配置して作成することが可能であることも実証した(Fig.5)。これは、前世代からのNanosheetを用いた各種IPを流用できることを意味しており、**回路設計上のメリットが大きい**。

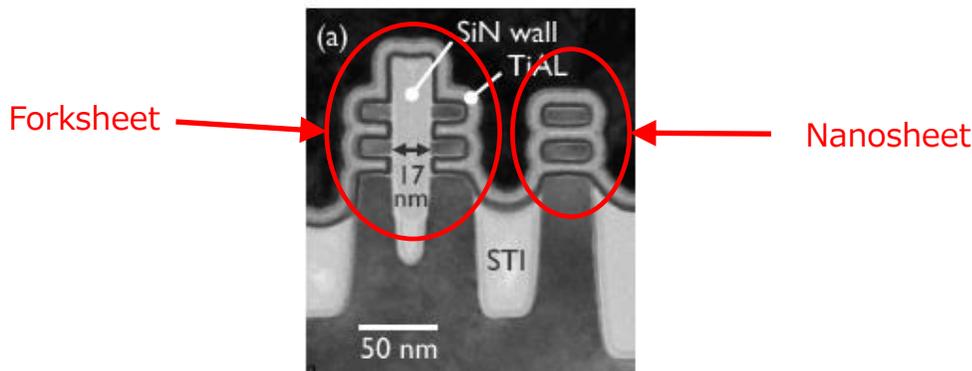


Fig. 5. X-section of Co-integration of Forksheet and Nanosheet.
©JSAP

4-1-1. フォークシート構造のN/Pトランジスタ技術 (4/4)

(5)本論文が注目される理由

Mooreの法則が終焉するかどうかは、長年、技術者間で議論されて久しいが、ベルギーの研究機関imecは、Mooreの法則の延命の為に各種施策について、シミュレーションベースでの提唱と、300mmラインでの試作を通じた実証を勢力的に行ってきた。 (imecには、Foundry大手のTSMC、Samsung、Intelが参加している。)

2nm世代Standard Cellの実レイアウトにおける微細化のために、Buried Power Rail (BPR)やBackside Power Delivery Network (BS-PDN)を提唱しているのはその一例である。

International Electron Device Meeting 2019 (IEDM2019)において、imecは、N-P距離を微細化でき、かつNanosheetと組み合わせることのできるForksheetsを立案し、シミュレーションベースでの効果を提唱していた。

今回のVLSI Symposiumでは、その構造をはじめ実際に試作して効果を実証した点で大いに注目に値する。

(6)実用化までの課題

今回は単体レベルのトランジスタ検証であったが、次のステップは、大規模ロジック回路・大規模SRAM回路での動作検証・歩留改善が必要となる。さらに、その次のステップは、プロセスコスト改善が課題になろう。

そもそも、本技術は、5-track Standard Cellの性能改善の為に必要な技術であるが、6-track cellならば必須では無い可能性もあるため、各FoundryはForksheets無しのNanosheetをもう一世代延命する可能性もある。

4-1-2. 2D材料チャネルMOSFET技術 (1/6)

T2-3

Advancing Monolayer 2D NMOS and PMOS Transistor Integration From Growth to van der Waals Interface Engineering for Ultimate CMOS Scaling

C. J. Dorow^{*}, K. P. O'Brien^{*}, C. H. Naylor, S. Lee, A. Penumatcha, A. Hsiao, T. Tronic, M. Christenson, K. Maxey, H. Zhu², A. Oni², U. S. Alaam, T. A. Gosavi, A. Sen Gupta, R. Bristol, S. Clendenning, M. Metz, U. E. Avci

Components Research and ²Quality and Reliability, Technology Development, Intel Corporation, Hillsboro, OR 97124, USA

^{*}Both authors contributed equally. Emails: chelsey.dorow@intel.com, kevin.p.obrien@intel.com

(1) 本論文が応用される分野、製品

2nm世代より先の世代のCMOS技術に適用可能性あり。その世代は、6G以降のMobile/高速通信、データセンター等のHPC、AI/Deep Learningに用いられる。

(2) 当該分野における技術ニーズ

2nmより先の世代においても、超高速・低消費電力・高集積度のMOSFETが必要とされている。それを実現するには、短チャネル効果を抑制するための極薄チャネルが必要であり、TMDなどの2D(2次元)材料をチャネルに適用したMOSFETを積層にした構造が、その候補となる。

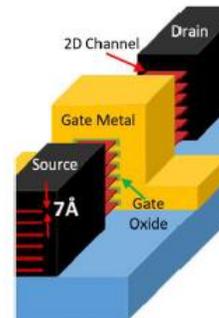


Fig. 1. Ultimate CMOS scaling with stacked 2D nano-sheets. ©JSAP

4-1-2. 2D材料チャネルMOSFET技術 (2/6)

(3)本論文の概要

2nm以降CMOSにおいては、短チャネル効果を抑制する観点から極薄チャネルが望まれているが、その候補として、2D材料(二次元材料)、その中でもTMD (Transition Metal Dichalcogenide; 遷移金属ダイカルコゲナイド)をチャネルに適用する事が研究されている。TMDとは、構成式MX₂からなり、「M」はMoやW等の遷移金属、「X」はSやSe等のカルコゲン原子である。また、TMD 2D材料を用いると、シリコンよりも高いチャネル移動度が得られることが期待されている。

インテルは、TMDによる2D材料チャネルがMOSFETトランジスタの究極のスケーリングを可能にし、 Mooreの法則に基づく半導体微細化を今後数十年にわたって可能にすることを示す研究成果を報告した。

厚さ1nm以下の単層TMDチャネルを用いて、CVD法、MBE法、シード成長などで製造したnおよびp-MOSFETの動作実証をした。新しいコンタクトメタルを用い、MBE法で形成した MoS₂チャネルNMOSデバイスは、ばらつきが少なく、既報では低いコンタクト抵抗(R_c) 0.4 kΩ・μm、小さいヒステリシス、77 mV/decの良好なサブスレッショルドスロープ(SS)値を得た。

CVD法で形成した WSe₂チャネルPMOSデバイスではCVD膜を用いた素子として最も良好なSS値89mV/decが得られた一方、オン電流はNMOSに及ばないことも分かった。

転写が不要な選択成長WS₂膜を用いたトランジスタでは、シード成長を用いたWS₂チャネルを用いたものとして最高のオン電流10 μA/μmを達成した。

4-1-2. 2D材料チャネルMOSFET技術 (3/6)

(4)技術的に特筆すべき点

以下の(a)-(c)の3点が特筆すべき点である。

(a) MoS2 NMOSについての性能改善

TMD 2D NMOSにおいては、MoS2を用いると良好な特性が得られることが知られているが、本論文では、MoS2のコンタクト抵抗に着目して、性能を改善させている。まず、CVD法MoS2よりもMBE法で形成したMoS2の方が低抵抗コンタクトになることを示した(Fig.5)。更に、Au(金)のコンタクトに替えて、“Novel Contact”を適用することで、大幅に低抵抗コンタクトとなることを示した。

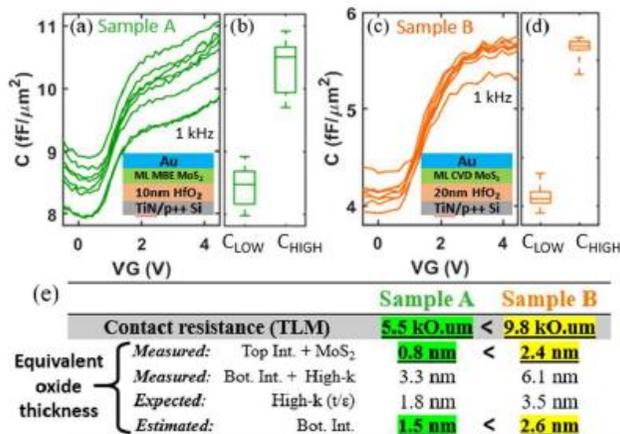


Fig. 5. (a) and (c) Measured capacitance vs gate voltage. (b) and (d) Extracted C_{LOW} and C_{HIGH} . (e) Sample with lower contact resistance also shows thinner electrical thickness at its interfaces. ©JSAP

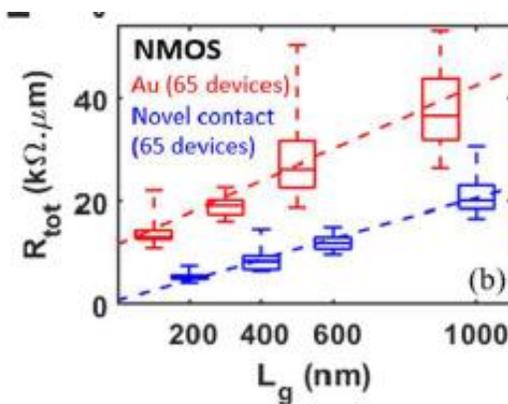


Fig. 7(b). Resistance vs gate length of NMOS. ©JSAP

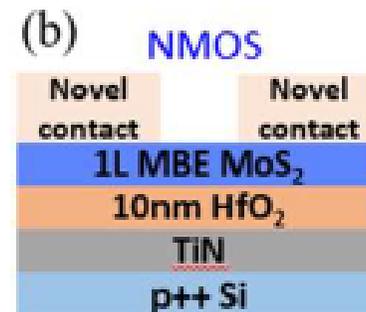


Fig. 6(b). Cross section of NMOS device. ©JSAP

4-1-2. 2D材料チャネルMOSFET技術 (4/6)

(b) TMD PMOSについての考察

一方、TMD 2D PMOSにおいては、業界でも中々高性能な微細MOSが出来ていない。現時点で、PMOSにはWSe₂が最も良いとされており、本論文では89mV/decという最も良好なSS値は達成した。一方、問題である高抵抗の原因が、WSe₂とPdメタルの間がショットキー接続になっていることを示した。Fig.6やFig.14 に示すようにNMOSと比較して、PMOSのオン電流が2桁低いことが見て取れる。

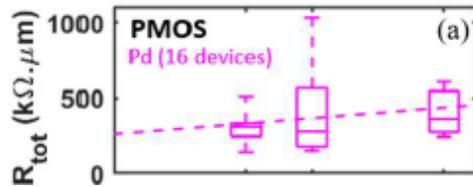


Fig. 7(a). Resistance vs gate length shows contact limited PMOS. ©JSAP

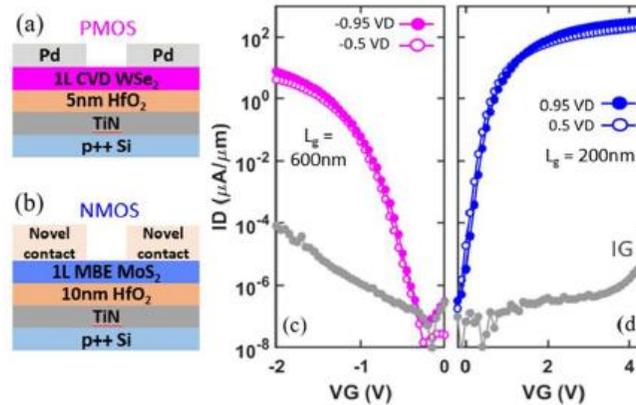


Fig. 6(a). Cross section of PMOS device.
Fig. 6. Id-Vg Curves for NMOS and PMOS.
©JSAP

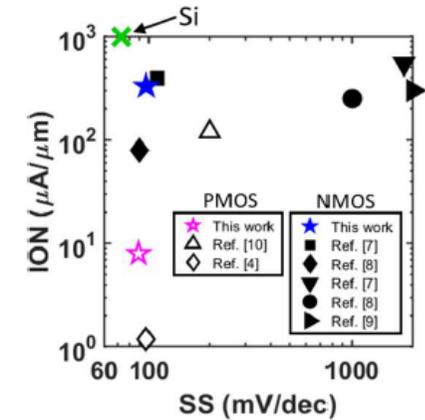


Fig.14 2D CMOS Benchmark.
©JSAP

4-1-2. 2D材料チャネルMOSFET技術 (5/6)

(c) TMD CMOSの可能性について言及

蒸着したメタルseedを酸化およびALDにより、Fig.12のようなNMOSとPMOSを形成した。PMOSでは1nm Hfを、NMOSでは1nm Alをseedとして用いた。これらのメタルシードの製法を用いれば、1種のTMD上にCMOSを作成できる可能性を示している。

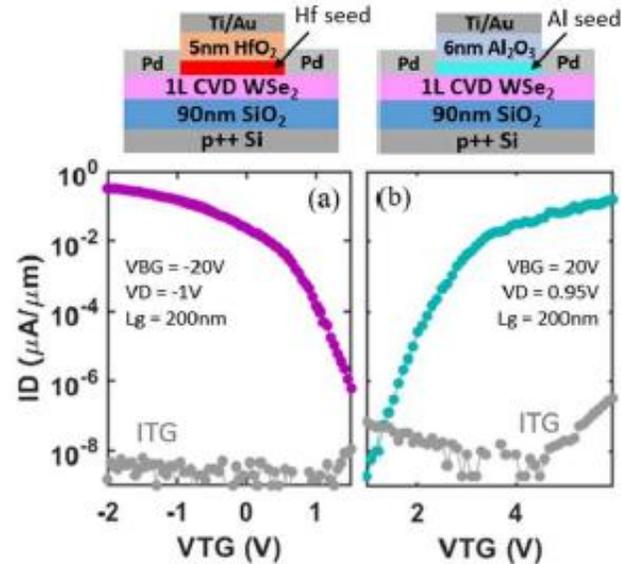


Fig. 12 Selectivity between NMOS and PMOS demonstrated by TG oxide formation with (a) Hf seeded ALD vs (b) Al seeded ALD. ©JSAP

4-1-2. 2D材料チャネルMOSFET技術 (6/6)

(5)本論文が注目される理由

2D材料は次世代MOSに適用される候補として近年、注目を集めているが、中々、Siに匹敵する性能がでてきておらず、プロセス改善が望まれてきた。今回の論文では、NMOSにおいてコンタクト抵抗を改善することで、オン抵抗を大幅に改善することを示せたことは評価に値する。一方、高抵抗のPMOSにおいても、Pd/MoS₂界面がショットキー接続になっていることを示し、今後の改善が待たれる。

(6)実用化までの課題

Intelのみならず、TSMCにおいても、2D材料は精力的に研究してきており、ポストSi材料として出現が待たれている。上述のFig.14に示す通り、NMOSは性能改善して来ているとは言え、まだSiの性能(Ion vs. SSカーブにおいて)に到達していない。また、PMOSについては、コンタクト部の材料系をどう変えていくかが大きな課題。

ロードマップにおいて、Si FinFETの次に、約3nm世代からSi Nanosheet (別名: NanoRibbonやMBCFET)を導入する。そのNanosheet構造に、いつ2D材料を組み合わせられるかで実用化時期が決まるが、上述の性能改善が最重要課題である。

4-1-3. 最先端3D NAND Flashメモリ (1/5)

T2-2

Highly Manufacturable 7th Generation 3D NAND Flash Memory with COP structure and Double Stack Process

Jun Hyoung Kim, Yongsik Yim, Joonsung Lim, Hyun Suk Kim, Eun Suk Cho, Chadong Yeo,
Woongseop Lee, Byungkwon You, Byoungil Lee, Minkyu Kang, Woojae Jang, Youngho Kwon,
Keehong Lee, Jaeduk Lee, Myeong-Cheol Kim, Jin-Yub Lee, Sunghoi Hur, Su Jin Ahn,
Hyeongsun Hong, Yu Gyun Shin, Hyoung-Sub Kim and Jaihyuk Song
Device Solutions, Samsung Electronics Co., Pyeongtaek 17786, Korea. E-mail) junh0903.kim@samsung.com

(1) 本論文が応用される分野、製品

最先端のConsumer SSD, Datacenter SSD

(2) 当該分野における技術ニーズ

Flashメモリベンダー各社は、3次元に積層したNAND Flashメモリを開発してきているが、多層数化により集積度をあげてきている。2019年には各社120層近くまで、各社層数をあげてきている。NAND Flashの記憶密度は、今まで、「2年で約2倍」のスピードで集積化が進んでおり、その為のプロセス技術が必要となる。

4-1-3. 最先端3D NAND Flashメモリ (2/5)

(3)本論文の概要

Samsungとしては、第7世代となる3D NANDを開発した。17X層(#)のワード線積層数を達成。COP構造(Cell Over Peripheral)を用いることで、tR 11%、tPROG 20%も改善した。第6世代と比較すると、セル体積スケーリングとCOP構造適用により、ビット密度は70%も増加できた。さらに、いくつかの新規プロセスの適用にも成功している。ダブルスタックプロセス、低ストレスタングステン、MSE(Multi Step Etch)、チャンネルホール側壁バッティングなどである。ダブルスタックプロセスは、チャンネルホールプロファイルを改善することにも寄与している。これらの結果、より良いセル動作特性が得られた。

(# 論文では17X層と一桁目を明示しなかったが、2021/6/8発表のSamsung Press room(※)では、17X層を176層と明記していた。)

※<https://news.Samsung.com/global/editorial-extraordinary-innovation-for-a-more-unforgettable-world-the-story-behind-samsungs-pioneering-v-nand-memory-solution>

TABLE I Feature Summary ©JSAP

	Device Generation	
	6th	7th
Bit Density	5Gb/mm ²	8.5Gb/mm ²
Stack	Single	Double
Peripheral	-	COP
WL W	-	Low Stress
Source Line Connection (Fig.7)	BCS	Side-wall butting
Erase Scheme	Body ERS	GIDL ERS

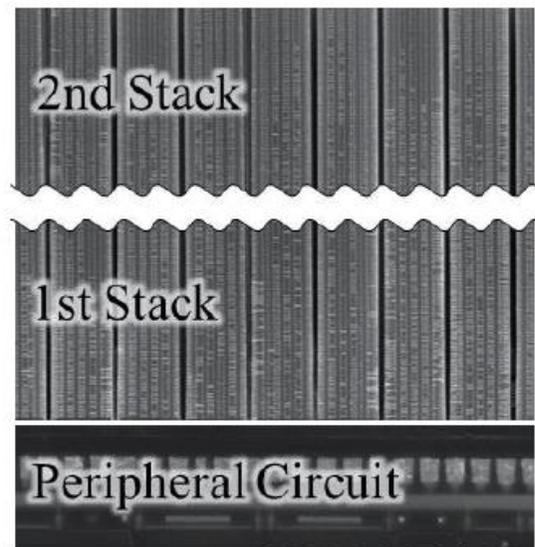


Fig. 1 Vertical SEM Image; COP(Cell Over Peripheral) and double stack process

©JSAP

4-1-3. 最先端3D NAND Flashメモリ (3/5)

(4)技術的に特筆すべき点

以下の(a)-(c)の3点が特筆すべき点である。

(a) Cell Volume ScalingとBit Density改善

3D NANDの微細化にとって、最も難易度が高いプロセスはコンタクトホール形成であるが、その難易度は[高さ2/ピッチ3]に比例する。Double Stack Processの適用により、難易度が大幅に低減した為、Unit Cell Volume (= BL pitch x Hole pitch x Unit Mold)の削減に貢献。更にCOP構造の適用もあって、Bit Densityは対前世代で+70%も改善した。

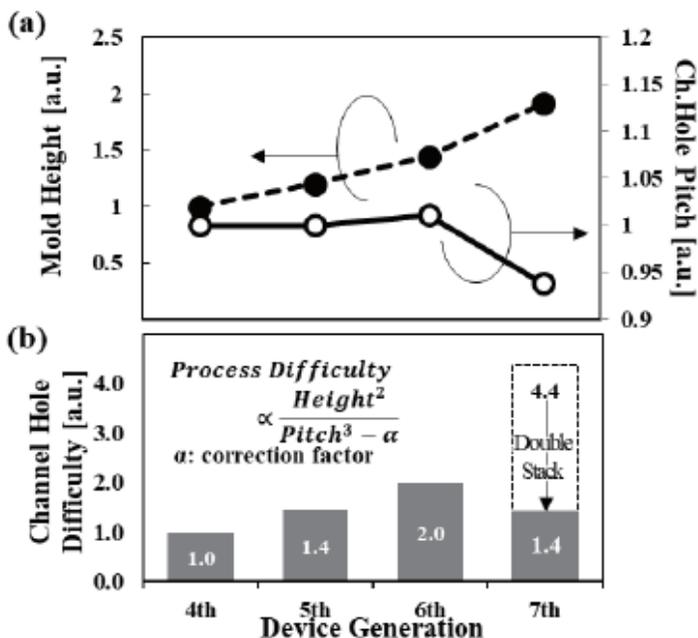


Fig. 2 (a) Mold height and channel hole pitch, (b) Channel hole difficulty by device generation

©JSAP

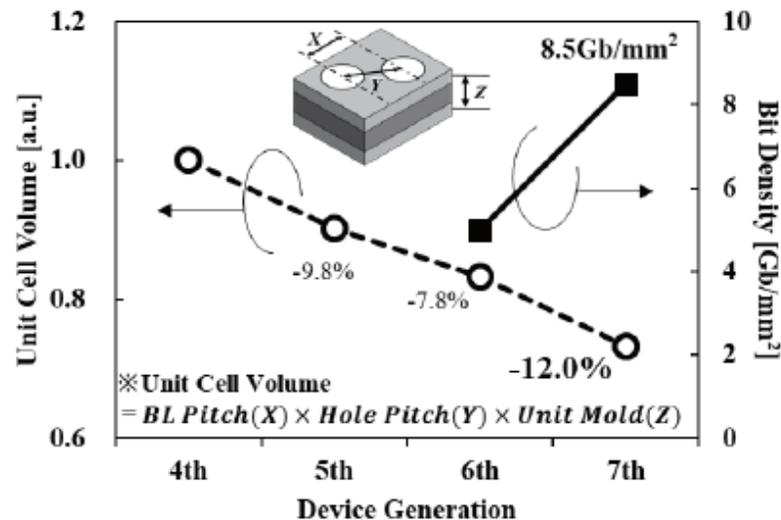


Fig. 3 Definition of unit cell volume and its trend over device generation. Cell volume reduction has been accelerated by double stack process and bit density has increased by 70%.

©JSAP

4-1-3. 最先端3D NAND Flashメモリ (4/5)

(b) Process上のBreakthrough (低ストレスW(タングステン), MSE(Multi Step Etch), チャンネルホール側壁バッティング) 高アスペクト比となると、セルブロックの倒れが生じやすくなるが、低ストレスのWに変更する事で、**倒れを改善した。**(Fig.5) また、MSEプロセスを適用することで、Wの間口が広がることで、**Wの膜質が改善した。**(Fig.6) さらに、従来はBCS (Body Contact Spacer)を用いていたが、今回、COP構造適用に伴い、Channel hole sidewall butting processを適用することで、**チャンネル電流伝導に対応している。**(Fig.7)

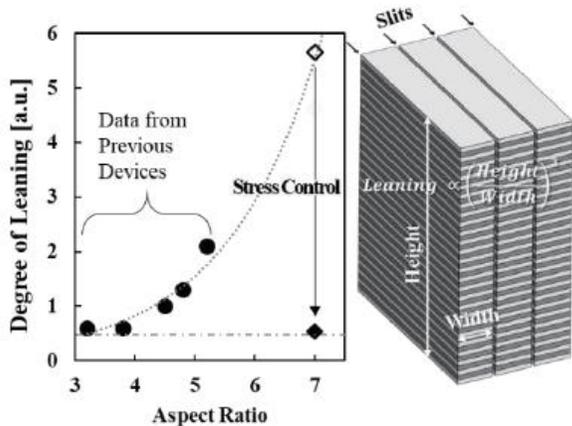


Fig. 5 Trend of leaning according to aspect ratio and its control with low stress W

©JSAP

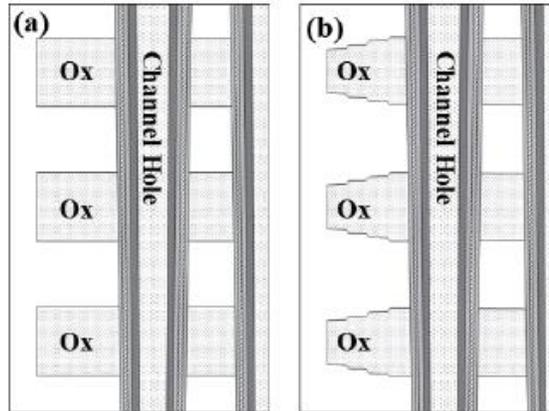


Fig. 6 Schematic comparison between (a) Normal WL and (b) WL with MSE process

©JSAP

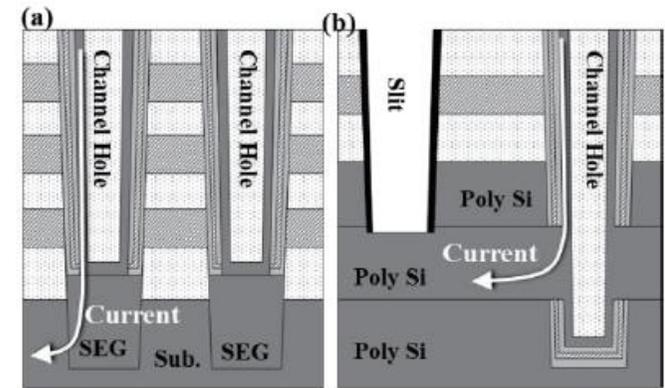


Fig. 7 Schematic comparison between (a) BCS (Body Contact Spacer) and (b) Channel hole sidewall butting process

©JSAP

4-1-3. 最先端3D NAND Flashメモリ (5/5)

(b) Cell Device特性の改善

COP構造の適用により、**tRが11%、tPROGが20%改善した。**(*) (Fig.8)

Cell Size Volumeがshrinkしたにも関わらず、Double Stack適用によるプロセス改善で**Cell to Cell結合が5%削減**できた。

さらに、ONO膜の最適化により、**Data retentionも改善**することができた。(Fig.9)

* tRは読み出しレイテンシ(メモリセルアレイからデータを出力バッファに転送するまでの時間)、tPROGは書き込みレイテンシ(入力バッファのデータをメモリセルアレイに書き込むまでの時間)である

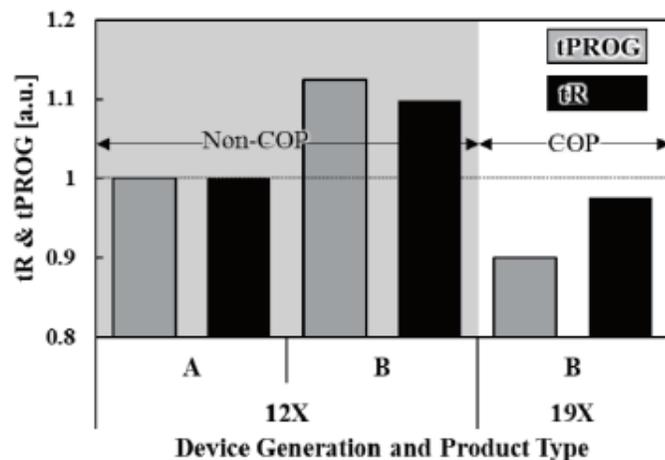


Fig. 8 The adoption of COP structure improved tR & tPROG by 11% and 20% respectively

©JSAP

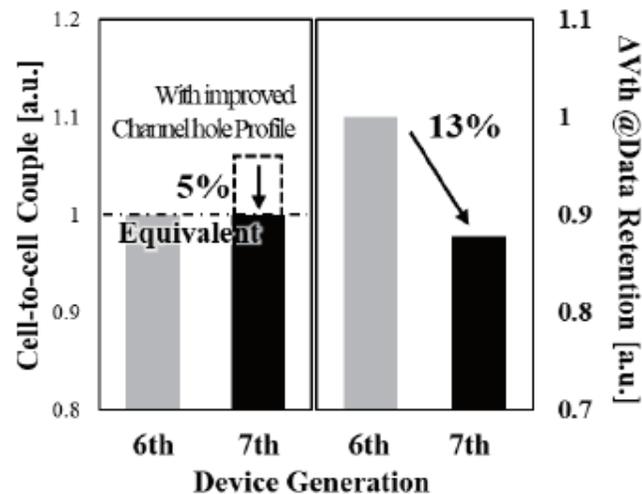


Fig. 9 Cell-to-cell coupling and data retention characteristics were improved by better channel hole profile and optimized ONO engineering

©JSAP

4-1-3. 最先端3D NAND Flashメモリ (6/5)

(5)本論文が注目される理由

半導体回路の国際学会（2021ISSCC：2021年2月開催）において、Samsung、SK-Hynix、Kioxia/WD連合から、170層以上の最先端 3D NAND回路についての発表はあった。一方で、プロセス・デバイスの本学会において、170層以上を有する最先端の3D NANDのプロセスの内容について、ある程度述べられている点は、非常に注目に値する。勿論、各社は量産プロセス詳細は隠したがる為、詳細までを把握することは出来ないが。（なお、SK-HynixやKioxia/WD連合からは今回本学会での発表は無し。）

(6)実用化までの課題

2021/6/8のSamsungの発表(※)では2021年下半期 Consumer SSDに、この第7世代 3D NANDを搭載するということであり、現在は、歩留改善と長期信頼性確認中であると思われる。また、今回新規に適用した各プロセスがどの程度コストUpに繋がっているかは現時点不明であるが、コスト削減は課題かもしれない。

※<https://news.Samsung.com/global/editorial-extraordinary-innovation-for-a-more-unforgettable-world-the-story-behind-samsungs-pioneering-v-nand-memory-solution>

4-2. Technology プレス発表されたその他注目論文 (1/4)

インパクトのある論文として紹介された残りの9件のリスト

論文番号	論文タイトル	著者、著者所属機関	論文概要
T2-1	Forksheet FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space	H. Mertens et al., imec, Belgium	<ul style="list-style-type: none"> ゲートアラウンド・ナノシートを利用したフォークシート構造のN/Pトランジスタを集積化 ゲート長22nmまで短チャネル効果を抑制(SS=66~68mV/dev)、チャネル形成後のウェットクリーニングを最適化し、トランジスタのオンオフ特性を改善 従来構造によるスケーリングの難点であるDual work function metal gatesの間隔を17nmまで縮小
T2-3	Advancing Monolayer 2D NMOS and PMOS Transistor Integration From Growth to van der Waals Interface Engineering for Ultimate CMOS Scaling	C. J. Dorow et al., Intel Corporation, USA	<ul style="list-style-type: none"> 厚さ1nm以下の単層TMD(Transition Metal Dichalcogenide)チャネルについて、CVD法、MBE法、シード成長などで製造したnおよびp-MOSFETの動作を実証 低いコンタクト抵抗(Rc)0.4kΩ・μm、小さいヒステリシス、77mV/dec CVD法で形成したWSe₂チャネルPMOSデバイスではCVD膜SS値89mV/dec
T15-2	First Highly Stacked Ge _{0.95} Si _{0.05} nGAAFETs with Record I _{on} =110μA(4100 μA/μm) at V _{OV} =V _{DS} =0.5V and High G _{m,max} =340μS(13000 μS/μm) at V _{DS} =0.5V by Wet Etching	Yi-Chun Liu et al., National Taiwan University, Taiwan	<ul style="list-style-type: none"> 8層Ge_{0.75}Si_{0.25}ナノシート構造と、7層Ge_{0.95}Si_{0.05}ナノワイヤー構造をH₂O₂ウェットエッチングで実現した V_{OV}=V_{DS}=0.5Vにおいて一層あたりI_{ON}=110μA(チャネルフットプリントあたり4100μA/μm)のレコードと、V_{DS}=0.5Vでの高いG_{m,max}=340μS(13000μS/μm)がGe/SeSi 3D nFETsで実証

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

4-2. Technology プレス発表されたその他注目論文 (2/4)

論文番号	論文タイトル	著者、著者所属機関	論文概要
T5-2	Dual Damascene BEOL Extendibility With Cu Reflow/Selective TaN And Co/Cu Composite	P. Bhosale et al., IBM Research, USA	<ul style="list-style-type: none"> デュアルダマシン(DD)を用いた配線接続工程の集積化が28nmピッチ以下まで微細化 2つの新しいプロセスフローとして、(1)信頼性を損なうことなくビア抵抗を低減できるようにTaNバリア膜の選択成膜を適用、(2)微細配線と幅広配線の性能向上のために金属材料技術を駆使した革新的な二種金属配線構造(dual metallization scheme)技術を開発 ハイ・パフォーマンス・コンピューティング(HPC)用途では、高アスペクト比の電源レールに二種金属配線(dual metallurgy)を用いることで最高性能に
JFS5-5	Buried nanomagnet realizing high-speed/low-variability silicon spin qubits: implementable in error-correctable large-scale quantum computers	S. Lizuka et al., AIST, Japan	<ul style="list-style-type: none"> シリコンスピン量子ビット動作を実現する埋め込み型微小磁石の集積方法 TCADに基づくシミュレーションにより、スピン量子ビットの極近傍に埋め込まれた微小磁石が生み出す大きな傾斜磁場により、高速な量子ゲート操作を実現し、量子ゲート操作の忠実度ばらつきを抑制 約10倍速いラビ振動(高速スピン反転)と、一定のプロセスのばらつきの条件下で99%以上の忠実度を予測 シリコンを用いた実用的な大規模集積量子コンピュータへの道を切り開く
T13-3	Fully on-chip MAC at 14nm enabled by accurate row-wise programming of PCM-based weights and parallel vector-transport in duration-format	P. Narayanan et al., IBM Research, USA	<ul style="list-style-type: none"> 14nmで試作したクロスバー型相変化メモリ(PCM)のテストチップ PCMをアナログ抵抗素子(ARES)として用いられ、クロスバー型PCMアレイのタイルが複数配列されてメッシュを形成して並列積和演算 誤差を3%未満に抑えるため、4つのPCMに行ごとの書き込み方式を採用 DNNのハードウェア実装を行い、画像認識、および再帰型長短期メモリ(LSTM)ネットワークによる誤差信号の伝搬に対する安定性を確認

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

4-2. Technology プレス発表されたその他注目論文（3/4）

論文番号	論文タイトル	著者、著者所属機関	論文概要
T2-4	First Demonstration of Atomic-Layer-Deposited BEOL-Compatible In ₂ O ₃ 3D Fin Transistors and Integrated Circuits: High Mobility of 113 cm ² /V·s, Maximum Drain Current of 2.5 mA/μm and Maximum Voltage Gain of 38 V/V in In ₂ O ₃ Inverter	Mengwei Si et al., Purdue University, USA	<ul style="list-style-type: none"> • BEOL工程で製造可能かつALDで成膜されたIn₂O₃ チャンネルを備えた3D Fin構造トランジスタ • In₂O₃ チャンネル平面構造トランジスタにおいて、チャンネル膜厚制御ならびに成膜後のアニール技術により113 cm²/Vの高いチャンネル移動度と、2.5 mA/μmの高いドレイン電流を実現 • ALD In₂O₃を用いて、最小駆動電圧(VDD) 0.5 V、最大電圧ゲイン38 V/Vのゼロゲートバイアス印加インバータ回路を実現
T2-5	3D stacked CIS compatible 40nm embedded STT-MRAM for buffer memory	M. Oka et al., Sony Semiconductor, Japan	<ul style="list-style-type: none"> • 世界初の積層型CMOSイメージセンサに適用可能な40nm embedded STTMRAM (Spin-Torque-Transfer Magnetic Random Access Memory) • CoFeBベースの縦型磁気トンネルジャンクション(p-MTJ)の最適化により、ウェーハ積層プロセスによって引き起こされる磁気特性の劣化を抑制 • -30℃における40ns以下の高速書き込み、105℃における10¹⁰回以上の書き換え耐性、85℃における1秒以上のデータ保持を達成
TFS2-6	Enabling Logic with Backside Connectivity via n-TSVs and its Potential as a Scaling Booster	A. Veloso et al., imec, Belgium	<ul style="list-style-type: none"> • ウェーハ薄化（STI酸化膜下のSi厚さを370nmから20nmまで薄化） W 充填ナノTSV（n-TSV）によって裏面接続を実現 • 信号と電源のネットワークを分離することを目的として、電源ネットワークをウェーハ裏面に移動することで、IRドロップの低減が期待 • nmos の移動度と駆動電流の向上（最大15%）、pmosの駆動電流ロスの減少（～3～10%） • 最終アニールのオプション選択が静電容量や信頼性の向上に有効

（資料）VLSIシンポジウム2021記者会見資料 ©JSAP

4-2. Technology プレス発表されたその他注目論文（4/4）

論文番号	論文タイトル	著者、著者所属機関	論文概要
T11-3	First Demonstration of Interface-Enhanced SAF Enabling 400oC-Robust 42 nm p-SOT-MTJ Cells with STT-Assisted Field-Free Switching and Composite Channels	Ya-Jui Tsou et al., Taiwan Semiconductor Research Institute(TSRI), Taiwan	<ul style="list-style-type: none"> • CMOSコンパチブルで400°Cの耐熱性を持つ42nmサイズの垂直スピン軌道トルク磁気トンネル接合（p-SOT-MTJ） • 合成反強磁性層(SAF)とイオンビームエッチング(IBE)技術により、トンネル磁気抵抗比(TMR) 130%を実現 • これまで報告された中では最も高い440°Cという耐熱性を持つSAF • スピン注入トルク（STT）アシストにより、外部補助磁界なしで確定的なSOT書き込みを実現
T6-3	Characterization of Fatigue and Its Recovery Behavior in Ferroelectric HfZrO	P. J. Liao et al., TSMC, Taiwan	<ul style="list-style-type: none"> • SILC（ストレス誘起リーク電流）測定に基づいて、HfZrO強誘電体キャパシタの分極疲労を調査 • 高電界ストレス下では分極の発現とSILCの増加には強い相関関係があり、低電界ストレス下での分極疲労はSILCを増加させることなく発生 • 低電界ストレスで発生する分極疲労を高電界ストレスの短周期サイクルを定期的に入れ込むことで疲労を回復させ、10^{12}回以上の書き換えを達成

（資料）VLSIシンポジウム2021記者会見資料 ©JSAP

5. Circuitsシンポジウム2021および注目論文の全体概要

● Circuits領域での世界の流れ

Circuits領域は情報の入力（センサ）・処理・記憶・通信、電源供給の5分野に大別される。AI/量子コンピューティングや5Gなどの注目アプリケーションに牽引されて、各分野において回路工夫、CMOSトランジスタの微細化、3次元実装などの技術工夫を組み合わせることにより大容量化・高速化・低消費電力化などの性能向上が確実に進展している。

● VLSIシンポジウムでの発表の動向

深層学習や強化学習などAI応用向けに計算量の多い演算を低消費電力で行うための専用チップ（アクセラレータ）や、量子コンピュータに向けた極低温で動作するチップに関する多くの発表が行われた。

● 将来的な研究開発の方向性

各分野に閉じこもってはいは技術限界が打破できないため、メモリ回路の中で機械学習演算を行うIn-memory computing、イメージセンサとAIチップの合体、プロセッサ上に集積した電源回路などを複数分野を融合した研究開発が活発に行われている。

● 注目論文全体概要

AI/量子コンピューティングや5Gなどを牽引役として、複数分野を融合した研究開発によりICチップの性能向上は確実に進展しており、今後も継続してICチップが豊かな未来社会実現に貢献するであろうことが伺われる論文が選定されている。

5-1. Circuits 特筆すべき論文

インパクトのある論文としてVLSI Circuitsシンポジウムの注目論文としての13件が紹介されており、その中で特筆すべき論文として以下の3件をピックアップした。それぞれの詳細について解説する。

※他の10件については、リストを掲載する。

論文番号	論文タイトル	著者所属機関
C3-1	A 32A 5V-Input, 94.2% Peak Efficiency High-Frequency Power Converter Module Featuring Package-Integrated Low-Voltage GaN NMOS Power Transistors	インテル
JFS1-3	A 7nm 0.46pJ/bit 20Gbps with BER 1E-25 Die-to-Die Link Using Minimum Intrinsic Auto Alignment and Noise-Immunity Encode	メディアテック
CFS1-2	CHIMERA: A 0.92 TOPS, 2.2 TOPS/W Edge AI Accelerator with 2 MByte On-Chip Foundry Resistive RAM for Efficient Training and Inference	スタンフォード大学

5-1-1. サーバ電源向けGaNパワーモジュール (1/5)

C3-1

A 32A 5V-Input, 94.2% Peak Efficiency High-Frequency Power Converter Module Featuring Package-Integrated Low-Voltage GaN NMOS Power Transistors

Nachiket Desai, Harish K. Krishnamurthy, William Lambert, Jingshu Yu, Han Wui Then, Nicolas Butzen, Sheldon Weng, Christopher Schaefer, N. Nidhi, Marko Radosavljevic, Johann Rode, Justin Sandford, Kaladhar Radhakrishnan, Krishnan Ravichandran, Bernhard Sell, James W. Tschanz and Vivek De

(1) 本論文が応用される分野、製品

データセンタのサーバ向けに5V入力1V出力の降圧型DC-DCコンバータを開発した。

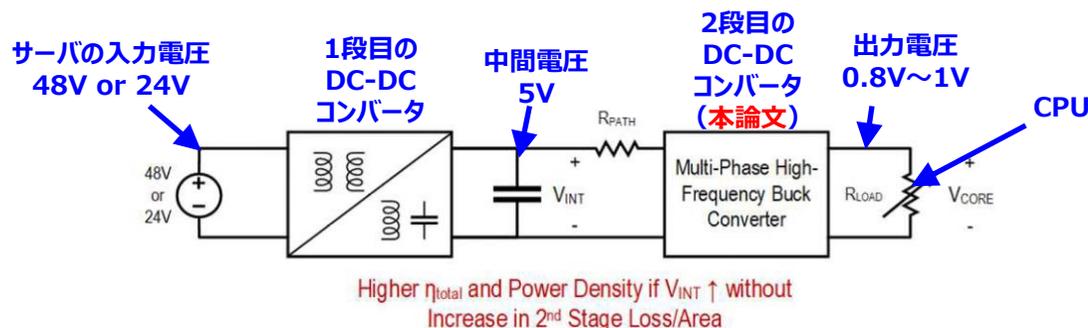


Figure 1: Challenges in high-density power delivery ©JSAP

(2) 当該分野における技術ニーズ

サーバのCPU向けに48V入力を1V, 100A出力に小型・高効率に降圧動作可能なDC-DCコンバータが求められている。

5-1-1. サーバ電源向けGaNパワーモジュール (2/5)

(3)本論文の概要

シリコンパワートランジスタよりも性能指数が5倍～10倍優れたGaNパワートランジスタを用いて5V入力、高周波数、高電流密度(9 A/mm²)の降圧型DC-DCコンバータを開発した。

4 mm角のパワーモジュールのパッケージにはGaNパワートランジスタとCMOSの駆動回路ICが集積されている。

40 nHのインダクタを用いて、このDC-DCコンバータを3 MHzのスイッチング周波数で動作させた場合、5V入力1V出力において、最大32 A出力、最高電力変換効率94.2 %を達成した。

(4)技術的に特筆すべき点

以下の(a)-(d)の4点が特筆すべき点である。

(a) 5V入力1V出力のDC-DCコンバータにおいて、実用的な大電流出力(最大 : 32 A)と高い電力変換効率(最高効率94.2 %)を実現している点。

(b) DC-DCコンバータのスイッチング周波数(f)が3.1 MHz - 10 MHzと高い点。

通常、f=100kHz - 1MHz程度。一般論として、DC-DCコンバータを小型化することが求められている。小型化するためにはfを高めることが有効。従来のシリコンのトランジスタは低速スイッチングのためfを高めることが容易ではないが、今回GaNトランジスタを使うことにより、f=10MHzを実現できた。

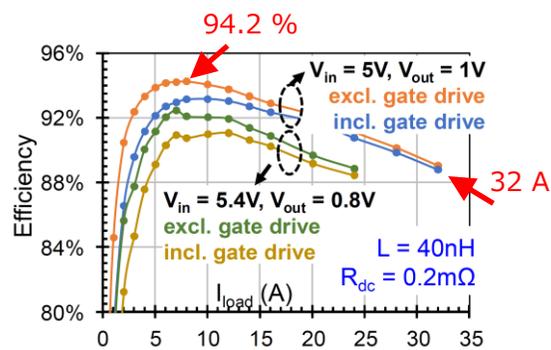


Figure 6: Measured efficiency (fsw = 3.1MHz) ©JSAP

Table 1: Performance comparison ©JSAP

	Schaefer, et al., ISSCC '19 [2]	Yan, et al., VLSI '20 [3]	Ke, et al., JSSC '18 [4]	Intel EN63A0QI [5]	EPC9036 [6]	This Work	
Architecture	4L FCML	Discrete GaN Buck	Discrete GaN Buck	MOSFET Buck	GaN Buck	LV GaN Buck with co-packaged CMOS Driver	
Technology	22nm CMOS	Discrete GaN + 0.18 μ m BCD	Discrete GaN + 0.35 μ m BCD	NA	E-mode GaN HEMT	LV GaN + 0.18 μ m CMOS	
V_{IN} (V)	5	5-24	3-40	2.5-6.6	12	3.6 - 5.4	
V_{OUT} (V)	0.8 - 1.8	1	1.2-5	0.6-6	1.2	0.8 - 1.8	
Max. I_{LOAD} (A)	10	1.2	1.2	12	32	32	22
fsw (MHz)	5	2.1	10	1.2	1	3.1	10
LOUT (nH)	10	NA	470	NA	250	40	10
Area (mm ²)	6.9	0.87 (Driver IC only)	0.86 (Driver IC only)	225	13.9	3.6	3.6
Die Area Current Density (A/mm ²)	1.45	NA	NA	NA	2.30	8.89	6.11
Efficiency @ Peak I_{LOAD}	89.4%	79%	73%	87%	86%	89%	83.5%
Peak Efficiency @ Conversion Ratio	88% @5:1 90% @3.7:1	90.2 @5:1 81% @12:1	85.5% @12:5	87% @5:1 90% @3.3:1	91.4 @12:1.2	94.2% @5:1	90.2% @3.6:1

5-1-1. サーバ電源向けGaNパワーモジュール (3/5)

(c) 大電力を扱うGaNトランジスタチップ(2 mm × 1.8 mm)とそれを駆動するシリコンのIC(1.3 mm × 1.3 mm)の2チップを4 mm角のパッケージ内に集積した**GaNパワーモジュールを実現**している点。シリコンのパワートランジスタに比べて、GaNトランジスタは約2桁高速にスイッチングするため、駆動回路とGaNトランジスタ間の距離を最短にして寄生インダクタンスを低減しないと誤動作する問題がある。世の中には駆動回路もGaNトランジスタチップに集積する手法とGaNトランジスタチップと駆動用シリコンICを近接配置する手法がある。今回は後者の手法を採用し、2チップを1つのパッケージ内に集積することにより駆動回路とGaNトランジスタ間の距離は0.3 mm以下を実現している。

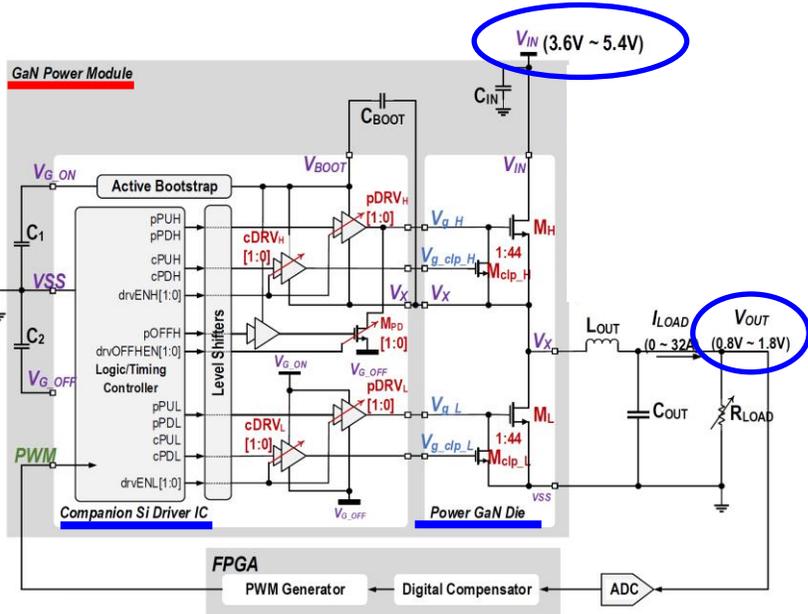


Figure 4: Block diagram of co-packaged GaN power module

©JSAP

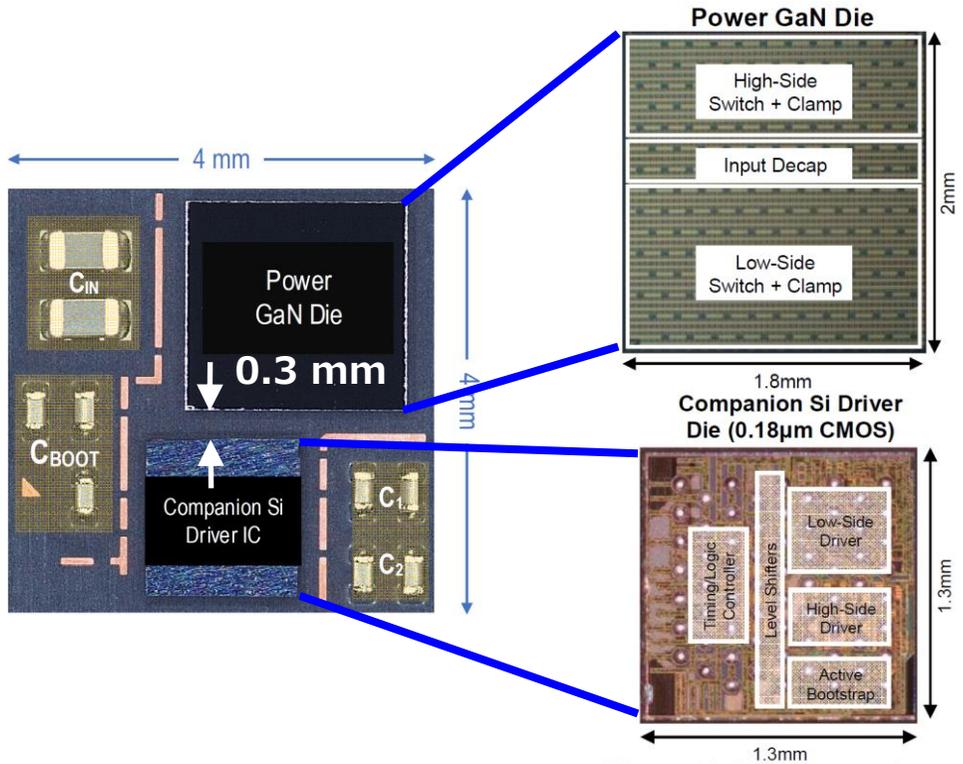


Figure 11: Die micrographs

©JSAP

5-1-1. サーバ電源向けGaNパワーモジュール (4/5)

(d) 5VというGaNトランジスタとしては低い電圧領域で、従来のシリコン微細FinFETトランジスタよりも**4倍～10倍高い性能**のGaNトランジスタを300 mmウェーハ上に実現した点。トランジスタの製造技術に関しては、本論文の引用文献[1]のIEDM 2019で既発表。市販のGaNトランジスタの耐圧は典型的には650 V, 100 Vであり、最低でも15 Vであった。

(根拠: <https://epc-co.com/epc/jp/%E8%A3%BD%E5%93%81/eGaNfET%E3%81%A8IC.aspx>)
 その結果として、GaNトランジスタの電流密度が9 A/mm²と高い点 (2 mm×1.8 mmのGaNチップで最大32 Aを流している)。従来のシリコンのパワートランジスタの電流密度は1 - 2 A/mm²程度である。

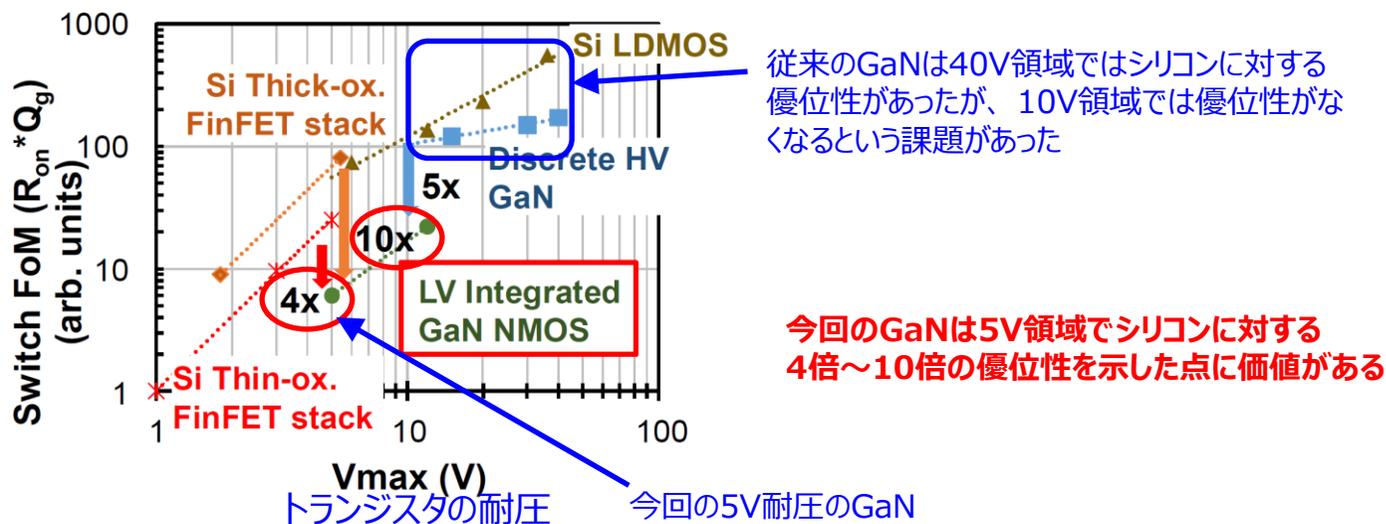


Figure 2: FoM comparison ©JSAP

5-1-1. サーバ電源向けGaNパワーモジュール（5/5）

(5)本論文が注目される理由

IntelはメニーコアのCPUの低消費電力化を目的としてDC-DCコンバータをチップ上に集積化するオンチップ電源回路（Integrated voltage regulator）の研究を約2005年頃から継続して行っている。本発表はその一環である。Intelは、これまで DC-DCコンバータのパワートランジスタとして微細CMOSを使ってきたが、今回、初めて、GaNを使った点が注目される。

しかも、GaNトランジスタをIntelが内製している点が注目される。これまで、パワートランジスタ業界（例えば、Infineon社、EPC社）と微細CMOSを扱うLSI業界（例えば、TSMC、Intel、Samsung）は別業界であり、GaNトランジスタを製造するのはパワートランジスタ業界の独壇場であった。しかし、今回、Intelが5Vという低電圧に特化して、オンチップ電源回路応用を見据えてGaNトランジスタをIntelが内製化した点は、今後の業界の枠を超えた競争を示唆しており、特に注目される。

(6)実用化までの課題

GaNトランジスタの一般的課題として、「高コスト」「低信頼性」があり、本論文のGaNトランジスタも研究レベルであるので、コストや信頼性は未知である。しかし、本論文のGaNトランジスタは300 mmウェーハ上に作成されているので、低コスト化できる可能性を示しており有望である。

5-1-2. 短距離・高速チップ間有線通信 (1/3)

JFS1-3

A 7nm 0.46pJ/bit 20Gbps with BER 1E-25 Die-to-Die Link Using Minimum Intrinsic Auto Alignment and Noise-Immunity Encode

Ying-Yu Hsu, Po-Chun Kuo, Chih-Lun Chuang, Po-Hao Chang, Hung-Hao Shen, Chen-Feng Chiang

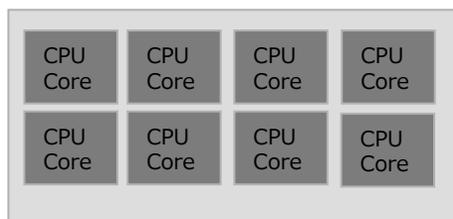
MediaTek Inc., Hsinchu Science Park, Hsinchu City, Taiwan, Ying-Yu.Hsu@mediatek.com

(1) 本論文が応用される分野、製品

Chipletのような複数チップが近接配置されたシステムにおけるチップ間の短距離・高速有線通信。

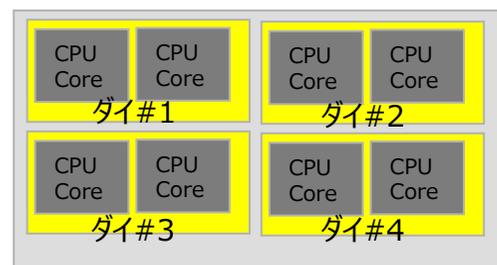
【Chiplet】 マルチチップモジュールで製造する方法。それぞれのチップは異なるプロセスで製造可能。一つの大面積チップの歩留まりと小面積の歩留まりでは、小面積の歩留まりの方が良いため、少面積チップを並べて使う方が安価であると言われている。Chipletのイメージを以下に示す。

従来



8個のCPUコアを1チップに入れる

Chiplet



2個のCPUコアが入ったチップを4個並べる

(2) 当該分野における技術ニーズ

CMOS技術の微細化が進むにつれ、チップの面積当たりの価格が高騰したため、複数の小面積チップを2次元アレーで並べて使うChipletの考え方が注目を集めている。しかし、Chipletを実現する上での最大の問題は高速・低消費電力・小面積のチップ間通信をいかに実現するかである。

5-1-2. 短距離・高速チップ間有線通信 (2/3)

(3) 本論文の概要

チップ間の極短距離通信向けに高密度、低ビットエラーレート、低消費電力の有線通信回路を開発した。TSMCの7 nm FinFET CMOSプロセスで有線通信回路ICを試作し、有線通信をTSMCの Chip-on-Wafer-on-Substrate(CoWoS)技術とIntegrated Fan-Out (InFO)パッケージ技術の両方で実証した。

(1) 周波数が低い4相のクロック信号の利用、(2) 自動遅延調整回路、(3) ノイズで誤動作しにくい新しい符号化などの技術工夫をすることにより、低消費電力で高速の有線通信を実現した。長さ1 mmの極短距離通信において、最大通信速度860 Gbps (20 Gbps/本×43本)、1ビット当たりのエネルギー0.46 pJ/bitの性能を達成した。通信密度は5.31 Tb/s/mm、2.25 Tb/s/mm²であり、従来例に比べ最も高い。

(4) 技術的に特筆すべき点

最先端の「**微細CMOSテクノロジー (7 nm)**」「**実装技術 (CoWoS, InFO)**」「**回路技術**」の3つを組み合わせることにより、短距離・高速チップ間有線通信として**世界最高性能を実現している点**が特筆すべき点である。

	This Work	[1] 2019 VLSI	[2] 2017 ISSCC	[3] 2018 ISSCC	[4] 2019 VLSI	[5] 2016 ISSCC	
Corp.	MediaTek	TSMC	Intel	AMD	Kandou	Kandou	
Technology	7nm FF	7nm FF	14nm FF	14nm FF	16nm FF	28nm	
Vsw(V)	0.8	0.3	0.85	low	-	-	
Channel	CoWoS 1mm (-3dB)	InFO 800um (-1dB)	CoWoS 500um	EMIB 1.1mm	MCM	MCM 30mm (-4.5dB)	MCM 12mm (-1.25dB)
uBump Pitch (um)	40.5	48.5	40	55	150	150	
Data Rate (Gbps/pin)	20	12.8	8	2	5.3	20.83	25
PHY Power Eff (pJ/bit)	0.46	0.5	0.56	1.2	2	1.02	0.94
Shoreline Throughput (Tbps/mm)	5.31	2.31	0.67	0.333	0.113*	0.42*	0.21*
Bandwidth Density (Tbps/mm ²)	2.25	1.57	0.8**	0.75**	0.056*	0.17*	0.25*

* Reference result scale up due to bump pitch 150um is normalized to 40um uBump pitch
** Reference showing result should be divide by 2 due to D2D (two die area)

微細CMOS

最先端の実装技術

高速動作

低エネルギー動作
(低消費電力)

高い通信密度

Fig. 8 Comparison to prior work. ©JSAP

5-1-2. 短距離・高速チップ間有線通信 (3/3)

(5)本論文が注目される理由

以下の2点が注目される。

(a) 「チップ間通信の高速・低消費電力化」という半導体産業の王道テーマにMediaTekが真正面から取り組んでいる点。台湾を代表するMediaTekは一昔前はスマートフォン向けのファブレスチップメーカーであったが、Intel, IBM, AMD, NVIDIAが得意とするクラウドのデータセンタ向けの高性能GPU, CPUへと事業範囲を拡大している点。今回発表した最大通信速度860 Gbps (20 Gbps/本×43本)という超高速有線通信技術は明らかにスマートフォンではなくデータセンタ向けである。データセンタ向けの高性能GPU, CPUではChipletが注目を集めており、本論文はChipletの課題を解決する重要技術を提供している。

(b) ファブレスのMediaTekとFoundryのTSMCが緊密に連携して開発を進めている点。論文の中にTSMCという用語が5回も登場し、MediaTekの論文であるがTSMCの宣伝であるかのような論文である。

(6)実用化までの課題

本論文は研究的な内容ではなく開発的な内容であるので、実用化に近い技術と思われる。

5-1-3. 学習可能なエッジAIチップ° (1/4)

CFS1-2

CHIMERA: A 0.92 TOPS, 2.2 TOPS/W Edge AI Accelerator with

2 MByte On-Chip Foundry Resistive RAM for Efficient Training and Inference

Massimo Giordano^{1†*}, Kartik Prabhu^{1†}, Kalhan Koul^{1†}, Robert M. Radway^{1†}, Albert Gural^{1†}, Rohan Doshi^{1†}, Zainab F. Khan¹, John W. Kustin¹, Timothy Liu¹, Gregorio B. Lopes¹, Victor Turbiner¹, Win-San Khwa², Yu-Der Chih², Meng-Fan Chang², GuénoLé Lallement^{1†}, Boris Murmann¹, Subhasish Mitra¹, Priyanka Raina¹

¹ Stanford University, CA, USA; ²TSMC, Hsinchu, Taiwan; [†]Equal Contribution; Email: mgiordan@stanford.edu

(1)本論文が応用される分野、製品

ユーザが取得したデータを元に、データをクラウドにアップロードせずに、ユーザの手元の例えばスマートフォンでAI学習させるためのエッジAIチップを開発した。

(2)当該分野における技術ニーズ

AI分野で注目されている深層学習は学習と推論の2つから構成される。推論に比べ学習には、膨大な量の計算とメモリが必要とされるためデータセンタにて計算処理が行われている。しかし、深層学習させるために、IoTで取得した様々なセンサデータすべてをデータセンタにアップロードすることは通信容量、消費電力、プライバシーの観点から現実的ではなく、ユーザの手元で深層学習を行う「エッジAI」が求められている。そこで、推論だけでなく学習を低消費電力で実現可能な「エッジAI」向け専用演算チップが強く求められている。

5-1-3. 学習可能なエッジAIチップ (2/4)

(3)本論文の概要

外付けのメモリを使わずにオンチップのメモリだけを使ってエッジAIの学習が可能な深層学習用チップを世界で初めて実現した。メモリとしては不揮発メモリであるReRAM（抵抗変化型メモリ）を2 MBオンチップに搭載した。本チップはピーク性能が0.92 TOPSであり、電力効率率は2.2 TOPS/Wである。本チップを6個接続して6倍大きな規模の深層学習にも拡張できる。ReRAMには書き込み回数に上限があるという欠点があるが、今回、ReRAMへの書き込み回数を1/283倍にする新しい学習アルゴリズムを提案し、この欠点を克服した。その結果、本チップを使えば10年間、3秒に1回の頻度でエッジAIの学習をし続けることができる。

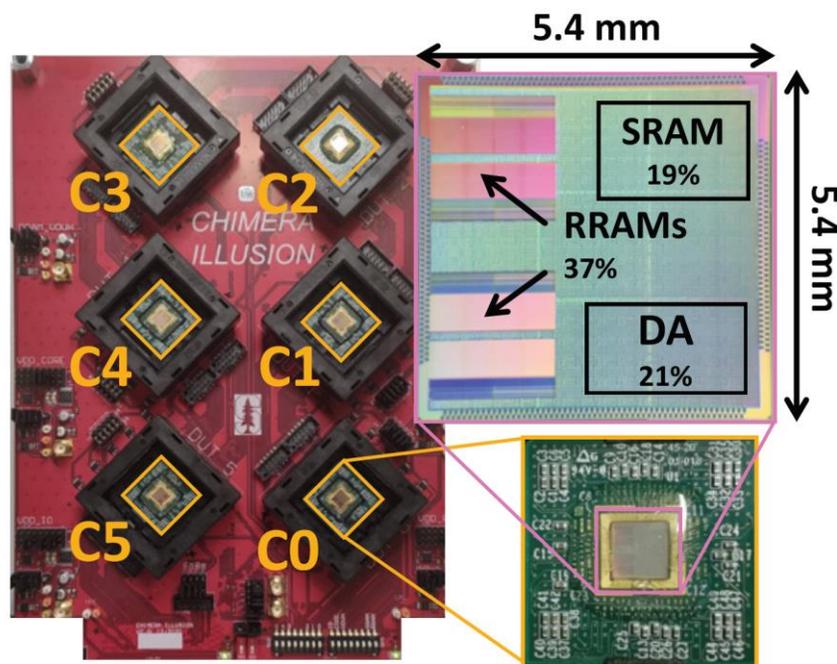


Fig. 14: Illusion System PCB, packaged die and micrograph. Die area: 29.16 mm². ©JSAP

5-1-3. 学習可能なエッジAIチップ (3/4)

(4) 技術的に特筆すべき点

外付けのメモリを使わずに**オンチップのメモリだけを使ってエッジAIの学習が可能**な深層学習用チップを世界で初めて実現した点。

オンチップメモリを使った学習に初めて成功した

外付けのメモリ(DRAM)を使って学習動作を行った先行例はあるが、DRAMを使うと性能低下と消費電力増大の課題があった

不揮発メモの使用により低消費電力動作を実現

オンチップメモリを使った先行例では**推論のみ**を行い、**学習**を行っていなかった

学習できるAIチップとしては世界最高性能

スマートフォンで利用可能な低消費電力動作に成功

	On-Chip RRAM		SRAM		Off-Chip DRAM		
	This work	Wang, VLSI'20	Zimmer, VLSI'19	Ueyoshi, ISSCC'18	Kang, ISSCC'20	Kim, ISSCC'19	Moons, ISSCC'17
Training	LRT	X	X	X	GANs	RL	X
Technology	40 nm	22 nm	16 nm	40 nm LP	65 nm	65 nm	28 nm
Area (mm ²)	29.2	10.8	6	121	32.4	16	1.9
Voltage (V)	1.1	0.6 - 1.1	0.42 - 1.2	0.77 - 1.1	0.7 - 1.1	0.67 - 1.1	1
Freq (MHz)	200	120	161 - 2000	75 - 330	25 - 200	5 - 200	200
RRAM	2 MB	3 MB	-	-	-	-	-
SRAM	0.5 MB	1 MB	752 KB	7.6 MB	576 KB	372 KB	144 KB
Precision	INT8,FP16	INT8	INT 8	INT1/4	FP8/16	INT 4/8/16	INT 1/16
TOPS	0.92	0.123	0.32 - 4	1.96 - 7.5	0.54 - 24.2	0.204	0.076
TOPS/W	2.2	0.96	0.2 - 9.1	0.6	1.8	0.55 - 2.16	0.2 - 10
Main Mem	8 pJ/B	30 pJ/B	-	15 pJ/B	50-200 pJ/B (estimated)		
Access	6 Gb/s	22.8 Gb/s	-	230 Gb/s	1-6 Gb/s (estimated)		
Power (mW)	126	128	30 - 4160	3300	58 - 647	2.4 - 196	19 - 62

LRT: Low Rank Training – GAN: Generative Adversarial Networks – RL: Reinforcement Learning

Fig. 15: Comparison versus state-of-the-art. ©JSAP

5-1-3. 学習可能なエッジAIチップ° (4/4)

(5)本論文が注目される理由

以下に示すICチップ設計分野におけるアメリカNo.1の教授陣と台湾No.1の教授とIC製造技術世界1位の台湾TSMCがドリームチームを結成して出した卓越した研究成果である点。

- (a) Boris Murmann : Stanford大学教授。**アナログ**回路、機械学習回路分野でアメリカNo.1。過去に国際学会ISSCCのプログラム委員長を務めた。
- (b) Subhasish Mitra : Stanford大学教授。**デジタル**回路、新デバイスを使った回路設計分野でアメリカNo.1。
- (c) Priyanka Raina : 2018年にMITにて博士号を取得し、現在、Stanford大学准教授。**AIチップ°**分野の新進気鋭の研究者。本論文のLast author (責任著者) である。
- (d) Philip Wong : Stanford大学教授。著者リストには含まれていないが、謝辞に記載にあるDARPA 3DSoCの参加研究者。IC**デバイス**研究分野でアメリカNo.1。2021 Symposia on VLSI Technology and CircuitsのExecutive Committee Chairを務める。数年前まで、台湾のTSMCにてVice President of Corporate Researchとして勤務。
- (e) Meng-Fan Chang : 台湾の国立精華大学教授。**メモリ**回路研究分野で世界的権威。現在、大学を離れTSMCにてDirectorとして勤務。

通常、**異分野のNo.1研究者が結集**することはないが、Stanford大学は見事にそれを成し遂げている。

また、一般に、TSMCは機密性の高いReRAMプロセスを大学には開示しないが、上記(d)(e)が**TSMCと太いコネクション**を有するため、ReRAMを使ったAIチップの設計・実証が可能になったと推測される。

(6)実用化までの課題

ReRAMは信頼性の課題があり研究レベルのメモリであり、世の中に広く普及するレベルではない。一方、TSMCは既に40nmのReRAMを少量であるが実用化している[1]ので、実用化の課題は少ないと思われる。

[1] <https://www.eenewsanalog.com/news/tsmc-offers-22nm-rram-taking-mram-16nm>

5-2. Circuits プレス発表されたその他注目論文（1/3）

インパクトのある論文として紹介された残りの10件のリスト

論文番号	論文タイトル	著者所属機関	論文概要
C11-1	A 28-GHz Phased-Array Relay Transceiver for 5G Network Using Vector-Summing Backscatter with 24-GHz Wireless Power and LO Transfer	東京工業大学	<ul style="list-style-type: none">• 5G中継用28GHz帯バッテリーレス無線機• ミリ波は強い指向性を持ち、また、伝搬距離も限られることから、基地局一台当たりでカバーできる通信範囲が狭くなる問題• 中継基地局を多数配置するために、24GHz帯無線電力伝送を併用するバッテリーレスでの5G規格準拠無線通信に成功
JFS3-4	Silicon Photonic Micro-Ring Modulator-based 4 x 112 Gb/s O-band WDM Transmitter with Ring Photocurrent-based Thermal Control in 28nm CMOS	インテル	<ul style="list-style-type: none">• 波長分割多重対応4 x 112Gb/s シリコンフォトニクス送信機• これまでマイクロリング変調器を用いた送信機の波長当たりのデータレートは50Gb/sであり、更なる高速化が要求されている• 4個のマイクロリング変調器の光電流を直接モニタし、動作温度をコントロールすることで各共振波長を安定させ、WDM送信機の高速化に成功した。
CFS1-3	OmniDRL: A 29.3 TFLOPS/W Deep Reinforcement Learning Processor with Dual-mode Weight Compression and On-chip Sparse Weight Transposer	KAIST	<ul style="list-style-type: none">• エネルギー効率に優れた深層強化学習(DRL)プロセッサ• 従来のDNNアクセラレータでは外部および内部のメモリアクセスが大きいことが、性能、消費エネルギーの課題• 指数平均デルタエンコーディング (EMDE) により指数圧縮率の向上 (1.6倍) とメモリアクセス電力の削減 (23.3%) を実現。さらに、スパース・ウェイト・トランスポザー (SWT) により圧縮された重みの転置をオンチップで行い、外部メモリへのアクセスを低減。4.18TFLOPSと29.3TFLOPS/Wを達成。

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

5-2. Circuits プレス発表されたその他注目論文 (2/3)

論文番号	論文タイトル	著者所属機関	論文概要
JFS2-5	HERMES Core - A 14nm CMOS and PCM-based In-Memory Compute Core using an array of 300ps/LSB Linearized CCO-based ADCs and local digital processing	IBM	<ul style="list-style-type: none"> 線形CCOを用いたAD変換器アレイおよびデジタル処理部を搭載した相変化メモリに基づくインメモリコンピューティングコア アナログデジタル変換 (ADC) などの周辺回路が電力やチップ面積のオーバーヘッド 提案コアは256x256乗算アレイ、周辺回路(ADC)、デジタル処理部を搭載。クロスバーアレイからの電流入力に対する発振周波数を線形に保つ補償回路を持つ新規設計のADCにより、1GHz動作可能な高精度の行列ベクトル乗算を実現。2つのコンピューティングコアの結合でMNISTおよびCIFAR-10データセットを用いた推論処理を実証し、エネルギー効率10.5 TOPS/W、面積性能1.59 TOPS/mm²を達成
C16-1	A 16Kb Antifuse One-Time-Programmable Memory in 5nm High-K Metal-Gate FinFET CMOS Featuring Bootstrap High Voltage Scheme, Read Endpoint Detection and Pseudo-Differential Sensing	TSMC	<ul style="list-style-type: none"> 5nm High-kメタルゲートFinFET製造プロセスによるアンチフューズ型16Kbワンタイムメモリ アンチフューズ素子の書き込みには高電圧が必要だが、最先端の微細CMOSではゲートリークの増加や絶縁耐圧の要因から動作マージンの確保が重要 世界初の5nm High-kメタルゲートFinFET製造プロセス向けのアンチフューズ型16Kbワンタイムメモリ (OTP) を開発。書き込みドライバの高耐圧トランジスタを絶縁破壊から保護するブートストラップ型の高電圧生成回路、読み出し時の電力および動作マージンを拡大させるエンドポイント検出回路、2素子を用いて弱いON状態となるディスタブ素子の影響を排除する疑似差動読み出し回路を実装することにより、SoCの出荷後プログラミングにおいて1ppb以下のエラーレートおよび、125℃環境下で10年間のデータ保持を実証
C18-1	A 19-GHz PLL with 20.3-fs Jitter	UCLA	<ul style="list-style-type: none"> 20.3fsジッタを実現する19GHz帯PLL 周波数が高くなるに従い必要とされる位相雑音(ジッタ)性能がより厳しくなることが近年大きな課題 28nm CMOS技術で作成した超低ジッタPLLを発表。両エッジサンプリング技術とタイミング制御技術を組み合わせることで、20.3fsのジッタ性能を実現し、250MHzの入力基準周波数に対して、12mWの消費電力で19GHzの発振周波数の出力を可能とした。

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

5-2. Circuits プレス発表されたその他注目論文 (3/3)

論文番号	論文タイトル	著者所属機関	論文概要
C15-1	A 10.0 ENOB, 6.2 fJ/conv.-step, 500 MS/s Ringamp-Based Pipelined-SAR ADC with Background Calibration and Dynamic Reference Regulation in 16nm CMOS	imec	<ul style="list-style-type: none"> 完全ダイナミック動作のパイプライン型SAR ADC 電源電圧や温度などの変動に対するロバスト性が課題 新規提案のダイナミック量子化器と狭帯域のディザ注入により、DACやステージ間ゲイン誤差、リングアンプのバイアス最適化に対応する高速なバックグラウンド較正を実現し、完全ダイナミック動作する基準電圧回路をチップ上に搭載。500MS/s動作時の消費電力は3.3mWであり、10.0ビットの有効分解能と75.5dBのSFDRを達成し、電力効率の指標であるWalden FoMは6.2fJ/変換ステップ達成。
JFS4-4	A 2.6 e-rms Low-Random-Noise, 116.2 mW Low-Power 2-Mp Global Shutter CMOS Image Sensor with Pixel-Level ADC and In-Pixel Memory	Samsung Electronics	<ul style="list-style-type: none"> Cu-to-Cuチップ接合技術による画素並列ADC搭載型CMOSイメージセンサ ADCの実装がイメージセンサの性能に大きく影響を与える。 高画質化と高速化を両立するために画素ごとにADCを搭載する構成を採用。光検出用のチップとADCを搭載したチップを画素並列に配置したCu PAD経由で高密度・高精度に接続。ビデオレートにおいて116.2mWの低消費電力と、2.6電子の低ノイズを実現し、最高で960フレーム/秒での動作可能。
C19-1	A Direct-Digitization Open-Loop Gyroscope Frontend with +/-8000°/s Full-Scale Range and Noise Floor of 0.0047°/s/√Hz	Robert Bosch	<ul style="list-style-type: none"> ジャイロスコープ用ワイドレンジフロントエンド回路 ジャイロセンサの出力は物理現象の特性上、信号レベルの範囲が大きいいため、計測回路にはワイドレンジ性が求められる。 回路方式の改善と最適化により、従来の同様の回路と比較して約4倍に相当する、±8000deg./sの大きな計測レンジを達成。そのほかの回路特性は従来回路と匹敵する性能を維持し、別の性能を犠牲にすることなく総合的な性能向上を達成。
C2-2	A Light Tolerant Neural Recording IC for Near-Infrared-Powered Free Floating Motes	ミシガン大学	<ul style="list-style-type: none"> 超小型ワイヤレス脳センサ向け神経信号測定チップ 駆動用の光がパッケージの隙間などからチップに届き、誤動作する問題 回路の構造と回路設計の工夫によって、光が当たる中でも安定して神経信号を計測し、処理したうえで送信する回路を実現。生体組織の限界である300μW/mm²の照射耐性を達成し、事実上、光による誤動作の問題を解消。加えて消費電力は38°Cで0.57μWであり、可変ゲイン制御と神経信号からの特徴抽出機能を搭載したスタンドアロン型の脳計測デバイスでは最小。

(資料) VLSIシンポジウム2021記者会見資料 ©JSAP

2-1-3. Hot Chips 33

目次 (1/2)

1. 開催概要

2. Hot Chips 33概要

3. プログラム

3-1. チュートリアル

3-2. 本会議 (基調講演およびセッション)

3-3. ポスター

4. Hot Chips 33発表の全体概要

4-1. 特筆すべき基調講演

4-1-1. K1: Builders of the Imaginary: From Artificial Intelligence to Artificial Architects
in the Era of SysMoore

4-1-2. K2: Skydio Autonomy Engine: Enabling the Next Generation of Autonomous Flight

4-2. 特筆すべきセッション論文

4-2-1. S1.3: The >5GHz next generation IBM Z processor chip

4-2-2. S2.2: Morpheus II: A RISC-V Security Extension for Protecting Vulnerable Software and Hardware

4-2-3. S4.3: Aquabolt-XL: Samsung HBM2-PIM with in-memory processing
for machine learning accelerators

4-2-4. S5.1: Accelerating ML Recommendation with over a Thousand RISC-V/Tensor Processors
on Esperanto's ET-SoC-1 Chip

4-2-5. S6.1: Graphcore Colossus Mk2 IPU

目次 (2/2)

4-3. 特筆すべきポスター論文

- 4-3-1. P04: Dynamic Neural Accelerator for Reconfigurable and Energy-efficient Neural Network Inference
- 4-3-2. P13: Edge Inference Engine for Deep & Random Sparse Neural Networks with 4-bit Cartesian-Product MAC Array and Pipelined Activation Aligner

【参考1】チュートリアルセッション発表概要

- 参考1-1. ML Performance and Real World Applications
- 参考1-2. Advanced Packaging

【参考2】その他注目論文

- 参考2-1. S1.1: Intel Alder Lake CPU Architectures
- 参考2-2. S1.2: AMD Next Generation “Zen 3” Core
- 参考2-3. S1.4: Next-Gen Intel Xeon CPU - Sapphire Rapids
- 参考2-4. S2.1: Mozart: Designing for Software Maturity and the Next Paradigm for Chip Architectures
- 参考2-5. S3.1: Arm Neoverse N2: Arm’s second-generation high performance infrastructure CPUs and system products
- 参考2-6. S3.2: NVIDIA DATA Center Processing Unit (DPU) Architecture
- 参考2-7. S5.3: Qualcomm Cloud AI 100: 12 TOPs/W Scalable, High Performance and Low Latency Deep Learning Inference Accelerator
- 参考2-8. S6.2: The Multi-Million Core, Multi-Wafer AI Cluster

【用語集】

1. 開催概要

1. 開催日時

2021年8月22日 - 24日 (3日間)

2. 開催形式

オンラインで開催

全てのチュートリアル、基調講演およびセッションはライブビデオストリームで配信

セッション完了後はオンデマンドで録画を視聴可能

ライブQ&Aや他の参加者とのディスカッションのためにSlackチャンネルが利用可能

※Slack登録者数は998名、シングルトラックの学会としては非常に大きな参加者数

3. 構成

下記3つのプログラムがある。

- ・チュートリアル：2チュートリアル
- ・本会議（基調講演およびセッション）：3基調講演、8セッション
- ・ポスター：18件



URL <https://hotchips.org/>

2. Hot Chips 33概要

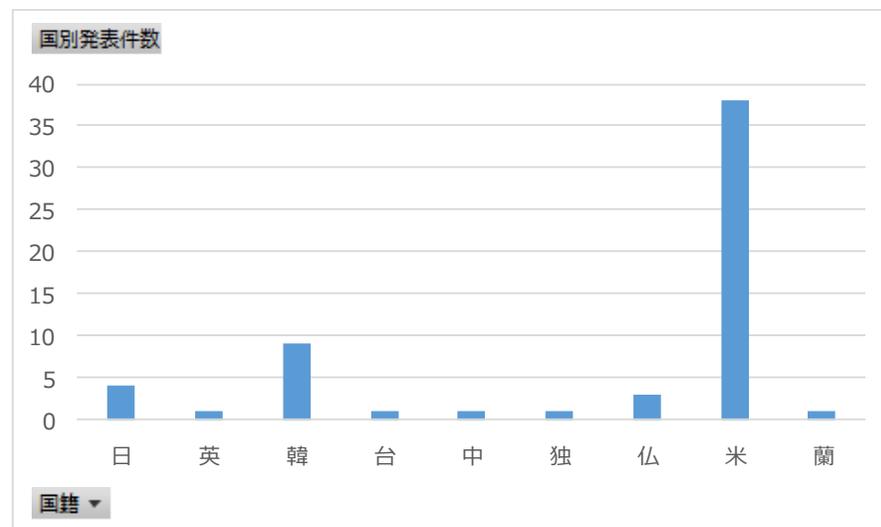
Hot Chipsは最先端のLSIチップに関するトップカンファレンスであり、マーケティング面・製品開発・システム技術が中心のカンファレンスとなっており、技術面・研究開発・回路技術中心のトップカンファレンスISSCCと棲み分けができています。IntelやAMDといったCPUのトップ企業や、NVIDIAやGoogleといったAIアクセラレータのトップ企業、また、GraphcoreやCerebrasといった新興企業が最先端の技術を発表する場となっている。

発表者所属機関別件数、国別件数を以下に示す。※チュートリアル・基調講演・セッション・ポスター全部で59件発表件数が最も多かったのが、Intelで7件、次いでSamsung Electronicsの6件、AMD、Google、KAISTの3件となっており、IDM(Intel, Samsungのような設計から製造・販売まで行うメーカー)からの発表が多数となった。その中で電気通信大学から**SoC要素技術**に関する2件、東京工業大学から**AIチップ**に関する1件のポスター発表があり日本の大学の存在感を高めている。

国別の発表件数では、米国が38件と突出しており、シリコンバレー開催の地の利と米国の技術力の高さを示している。セッションへの投稿論文は85本あり、うち27本が採択され、採択率は31.7%となり今年のVLSIシンポジウムよりも低くなっており、Hot Chipsの論文審査が厳格であり、採択論文の質が高く保たれていると言える。

#	発表者所属機関	発表件数
1	Intel	7
2	Samsung Electronics	6
3	AMD	3
4	Google	3
5	KAIST	3
6	Graphcore	2
7	LightOn	2
8	NVIDIA	2
9	電気通信大学	2

※2件以上の発表があった所属機関を記載



3. プログラム

3-1. チュートリアル

開催初日の22日に以下の2つのチュートリアルセッションが開催された。
チュートリアルセッションの概要を以下に、個々の発表概要を【参考1】に示す。

■ Tutorial 1: ML Performance and Real World Applications

Abstract	概要
Machine learning is a rich, varied, and rapidly evolving field. This tutorial will explore the applications, performance characteristics, and key challenges of many different unique workloads across training and inference. In particular, we will focus on hardware/software co-optimization for the industry-standard MLPerf™ benchmarks and selected applications and considerations at prominent cloud players.	機械学習は、豊富で多様性に富み、急速に発展している分野である。このチュートリアルでは、さまざまなワークロードのアプリケーション、パフォーマンス特性、および主な課題について、トレーニングや推論全体にわたって説明する。特に、業界標準のMLPerf™ベンチマークのハードウェアとソフトウェアの同時最適化、および主要なクラウドプレイヤーにおける特定のアプリケーションと考慮事項に焦点を当てる。

■ Tutorial 2: Advanced Packaging

Abstract	概要
This tutorial will discuss advanced 3D packaging technologies that enable performance and density improvements. Descriptions of the technologies and how they are used in cutting edge applications will be made by industry leaders in packaging and chip design.	このチュートリアルでは、パフォーマンスと密度を向上させる高度な3Dパッケージング技術について説明する。最先端のアプリケーションでこれらの技術がどのように使用されているかについては、パッケージングおよびチップ設計の業界リーダーが説明する。

(資料) <https://hotchips.org/>

3-2. 本会議（基調講演およびセッション）（1/4）

本会議は、23日、24日に実施され、3つの基調講演と8つのセッションが開催された。

■ 基調講演 ※赤字の講演は特筆すべき基調講演として後述

講演番号	Title	Presenters
K1	Builders of the Imaginary: From Artificial Intelligence to Artificial Architects in the Era of SysMoore	Aart de Geus, CEO, Synopsys
K2	Skydio Autonomy Engine: Enabling the Next Generation of Autonomous Flight	Abraham Bachrach, CTO, Skydio
K3	Architectural Challenges: AI Chips, Decision Support and High Performance Computing	Dimitri Kusnezov, Deputy Under Secretary for AI and Technology, Department of Energy

■ セッション

Session Title	Abstract
CPUs	Intel, AMD, IBMが各社のハイエンドCPUを発表
Academic Spinout Chips	Wisconsin-Madison大学、ミシガン大学から大学発のチップを発表
Infrastructure and Data Processors	Arm, NVIDIA, IntelがCPUを補助してセキュリティや処理性能を補助するプロセッサを発表
Enabling Technologies	Infineonが自動車システム用のヘテロジニアスコンピューティング、EdgeQが無線アクセス用RISC-V、SamsungがHBM2-PIM(Process-In-Memory)を発表
ML Inference for the Cloud	Esperanto Technologies, Enflame Technology, Qualcommがクラウド向け推論チップを発表
ML and Computation Platforms	Graphcore, Cerebras, SambaNova, D.E.Shaw Researchが各社のAIチップを発表
Graphics and Video	Intel, AMD, Google, XilinxがGPU・VCU(Video Coding Unit)等を発表
New Technologies	Mojo VisionがARコンタクトレンズ、Samsungが全方向オートフォーカスレンズ、香川大学が指先識別能力を超えるナノ触覚センサーチップ、IonQが量子コンピュータを発表

(資料) <https://hotchips.org/>

3-2. 本会議（基調講演およびセッション）（2/4）

■ CPUs

論文番号	Title	Presenters
S1.1	Intel Alder Lake CPU Architectures	Efraim Rotem, Intel
S1.2	AMD Next Generation "Zen 3" Core	Mark Evers, AMD
S1.3	The >5GHz next generation IBM Z processor chip	Christian Jacobi, IBM
S1.4	Next-Gen Intel Xeon CPU - Sapphire Rapids	Arijit Biswas and Sailesh Kottapalli, Intel

■ Academic Spinout Chips

論文番号	Title	Presenters
S2.1	Mozart: Designing for Software Maturity and the Next Paradigm for Chip Architectures	Karu Sankaralingam, University of Wisconsin- Madison
S2.2	Morpheus II: A RISC-V Security Extension for Protecting Vulnerable Software and Hardware	Todd Austin, University of Michigan

■ Infrastructure and Data Processors

論文番号	Title	Presenters
S3.1	Arm Neoverse N2: Arm's second-generation high performance infrastructure CPUs and system products	Andrea Pellegrini, ARM
S3.2	NVIDIA DATA Center Processing Unit (DPU) Architecture	Idan Burstein, NVIDIA
S3.3	Intel's Hyperscale-Ready SmartNIC for Infrastructure Processing	Bradley Bures, Intel

※赤字の発表は特筆すべきセッション論文として後述、青字の発表はその他注目論文として後述

(資料) <https://hotchips.org/>

3-2. 本会議（基調講演およびセッション）（3/4）

■ Enabling Technologies

論文番号	Title	Presenters
S4.1	Heterogeneous computing to enable the highest level of safety in automotive systems	Ramanujan Venkatadri, Infineon
S4.2	Architecting an Open RISC-V 5G and AI SoC for Next Generation 5G Open Radio Access Network	Sriram Rajagopal, EdgeQ
S4.3	Aquabolt-XL: Samsung HBM2-PIM with in-memory processing for machine learning accelerators	Jin Hyun Kim, Samsung Electronics

■ ML Inference for the Cloud

論文番号	Title	Presenters
S5.1	Accelerating ML Recommendation with over a Thousand RISC-V/Tensor Processors on Esperanto's ET-SoC-1 Chip	David Ditzel, Esperanto Technologies
S5.2	AI Compute Chip from Enflame	Ryan Liu and Chuang Feng, Enflame Technology
S5.3	Qualcomm Cloud AI 100: 12 TOPs/W Scalable, High Performance and Low Latency Deep Learning Inference Accelerator	Karam Chatha, Qualcomm Inc

■ ML and Computation Platforms

論文番号	Title	Presenters
S6.1	Graphcore Colossus Mk2 IPU	Simon Knowles, Graphcore
S6.2	The Multi-Million Core, Multi-Wafer AI Cluster	Sean Lie, Cerebras Systems
S6.3	SambaNova SN10 RDU: Accelerating Software 2.0 with Dataflow	Raghu Prabhakar and Sumti Jairath, SambaNova Systems, Inc
S6.4	The Anton 3 ASIC: a Fire-Breathing Monster for Molecular Dynamics Simulations	J. Adam Butts, D.E. Shaw Research

3-2. 本会議（基調講演およびセッション）（4/4）

■ Graphics and Video

論文番号	Title	Presenters
S7.1	Intel's Ponte Vecchio GPU Architecture	David Blythe, Intel
S7.2	AMD RDNA(TM) 2 Graphics Architecture	Andrew Pomianowski, AMD
S7.3	Google's Video Coding Unit (VCU) Accelerator	Aki Kuusela and Clint Smullen, Google
S7.4	Xilinx 7nm Edge Processors	Juanjo Noguera, Xilinx

■ New Technologies

論文番号	Title	Presenters
S8.1	Mojo Lens - AR Contact Lenses for Real People	Michael Wiemer and Renaldi Winoto, Mojo Vision
S8.2	World Largest Mobile Image Sensor with All Directional Phase Detection Auto Focus Function	Sukki Yoon, Samsung Electronics
S8.3	New Value Creation by Nano-Tactile Sensor Chip Exceeding our Fingertip Discrimination Ability	Hidekuni Takao, Kagawa University
S8.4	The IonQ Trapped Ion Quantum Computer Architecture	Christopher Monroe, IonQ, Inc

(資料) <https://hotchips.org/>

3-3. ポスター (1/2)

以下の18件が発表された

論文番号	Title	Authors & Affiliation
P01	OmniDRL: An Energy-Efficient Mobile Deep Reinforcement Learning Accelerators with Dual-mode Weight Compression and Direct Processing of Compressed Data	Juhyoung Lee; Korea Advanced Institute of Science and Technology
P02	Exynos 1080 high-performance, low-power CPU and GPU with AMIGO	Taehee Lee; Samsung
P03	An Energy-efficient Floating-Point DNN Processor using Heterogeneous Computing Architecture with Exponent-Computing-in-Memory	Juhyoung Lee; Korea Advanced Institute of Science and Technology
P04	Dynamic Neural Accelerator for Reconfigurable and Energy-efficient Neural Network Inference	Sakyasingha Dasgupta; EdgeCortix
P05	SM6: A 16nm System-on-Chip for Accurate and Noise-Robust Attention-Based NLP Applications	Thierry Tambe; Harvard
P06	ENIAD: A Reconfigurable Near-data Processing Architecture for Web-Scale AI-enriched Big Data Service	Jialiang Zhang; U Penn
P07	A Plug-and-Play Universal Photonic Processor for Quantum Information Processing	Caterina Taballione; QuiX
P08	Industry's First 7.2 Gbps 512 GB DDR5 Memory Module with 8-Stacked DRAMs: A Promising Memory Solution for Next-Gen Servers	Sung Joo Park; Samsung
P09	LightOn Optical Processing Unit: Scaling-up AI and HPC with a Non von Neumann co-processor	Laurent Daudet; LightOn

※赤字の発表は特筆すべきポスター論文として後述

(資料) <https://hotchips.org/>

3-3. ポスター (2/2)

論文番号	Title	Authors & Affiliation
P10	System-on-Chip Implementation of Trusted Execution Environment with Heterogeneous Architecture	Trong-Thuc Hoang; University of Electro-Communications
P11	A CORDIC-based Trigonometric Hardware Accelerator with Custom Instruction in 32-bit RISC-V System-on-Chip	Khai-Duy Nguyen; University of Electro-Communications
P12	A photonic neural network using < 1 photon per scalar multiplication	Tianyu Wang; Cornell
P13	Edge Inference Engine for Deep & Random Sparse Neural Networks with 4-bit Cartesian-Product MAC Array and Pipelined Activation Aligner	Kota Ando; Tokyo Institute of Technology
P14	Photonic co-processors in HPC: using LightOn OPUs for Randomized Numerical Linear Algebra	Daniel Hesslow; LightOn
P15	Elpis: High Performance Low Power Controller for Data Center SSDs	Seungwon Lee; Samsung
P16	SOT-MRAM – Third generation MRAM memory opens new opportunities	Jean-Pierre Nozières; Antaios
P17	PNNPU: A Fast and Efficient 3D Point Cloud-based Neural Network Processor with Block-based Point Processing for Regular DRAM Access	Sangjin Kim; Korea Advanced Institute of Science and Technology
P18	Samsung NPU: An AI accelerator and SDK for flagship mobile AP	Jun-Seok Park; Samsung

(資料) <https://hotchips.org/>

4. Hot Chips 33発表の全体概要

● セッション

主に企業主体の発表となっている。学会名称が示唆する通り、特にサーバ、HPC寄りの発表が多い。

● 発表論文の特徴

今年は、センサー系など、従来のHot Chipsではなかった新規技術系の発表が数件見られるのが特徴であった。

● 特筆すべき論文

AI時代に合わせて従来の汎用機アーキテクチャに多くの工夫を施した発表、RISC-Vアーキテクチャの工夫でセキュリティを作り込むチップの安全性を向上させる発表、チャレンジング・主流とは違う独自の取り組みの発表、注目を集めていた製品の初めての技術詳細発表の5件を特筆すべきセッション論文として、また、ポスター論文からは日本に関連する発表（日本に拠点を置く珍しいAIチップスタートアップの発表、日本の大学として唯一AIチップ分野で成果を出しているグループの発表）2件を特筆すべきポスター論文としてピックアップして紹介する。

特筆すべきとまではいかないが興味深い発表は参考として概要を紹介する。

4-1. 特筆すべき基調講演

特筆すべき基調講演としてAI時代におけるEDA企業としての役割・ビジョンがよく分かる講演としてSynopsysの講演、今後のAIのさらなるチャレンジを感じさせる自律飛行ドローンに関するSkydioの講演をピックアップした。それぞれのポイントについて解説する。

講演番号	Title	Presenters
K1	Builders of the Imaginary: From Artificial Intelligence to Artificial Architects in the Era of SysMoore	Aart de Geus, CEO, Synopsys
K2	Skydio Autonomy Engine: Enabling the Next Generation of Autonomous Flight	Abraham Bachrach, CTO, Skydio

4-1-1. K1: Builders of the Imaginary: From Artificial Intelligence to Artificial Architects in the Era of SysMoore (Aart de Geus, CEO, Synopsys)

(1)本講演の概要

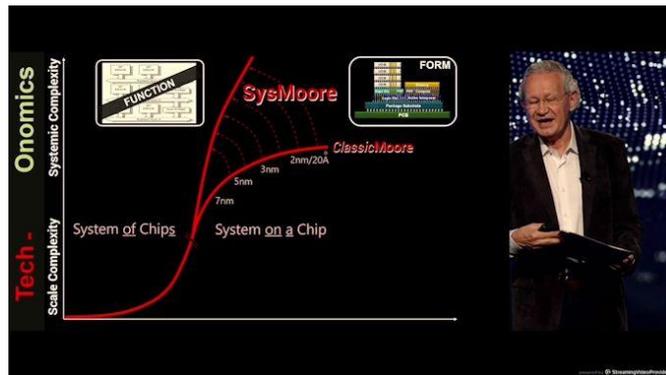
- システムレベルの技術でムーアの法則並の等比級数的成長を担うSysMooreの考え方を提言し、その中で設計ツール会社であるSynopsysの役割を強調。
- TechnologyのPushとEconomyのPullによりAI時代が爆発的に進展している。果たしてAIがチップを設計できるかという問いに対して、以下のように発表。
 - ✓ 時代はようやくCADからEDAに移行
 - ✓ AIの高い設計空間探索能力により、次第に現実のものになりつつある。

(2)本講演が注目される理由

- EDA業界の著名人によるAI時代そのものを俯瞰する非常にビジョナリーでクオリティの高い好講演

※下記サイトに詳しい情報がある。

<https://www.anandtech.com/show/16902/hot-chips-2021-day-1-keynote-1-designing-chips-with-ai-synopsys>



SysMoore:最先端の集積回路設計の説明用語。ムーアの法則のスケールの複雑さとハイパーコンバーズド統合のシステムの複雑さを組み合わせたもの。Synopsysが作った用語。

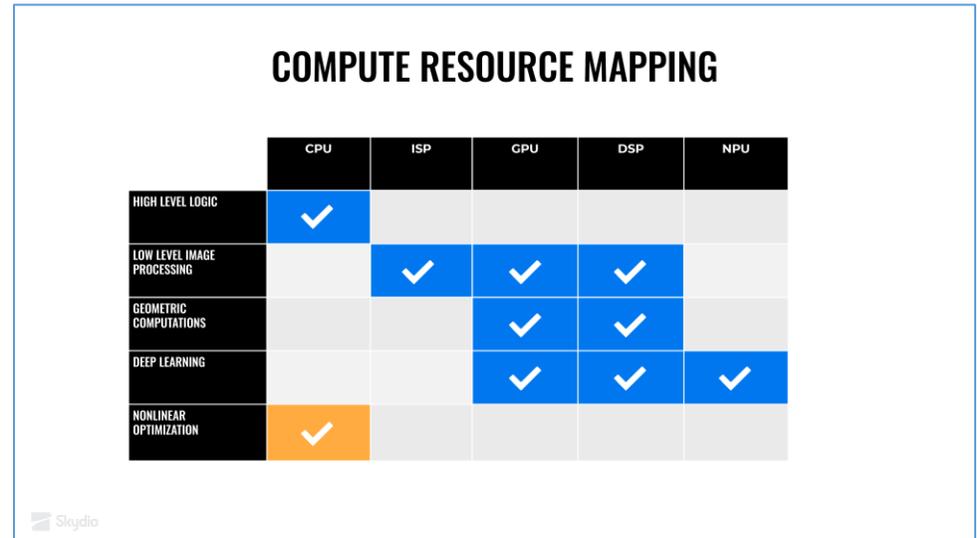
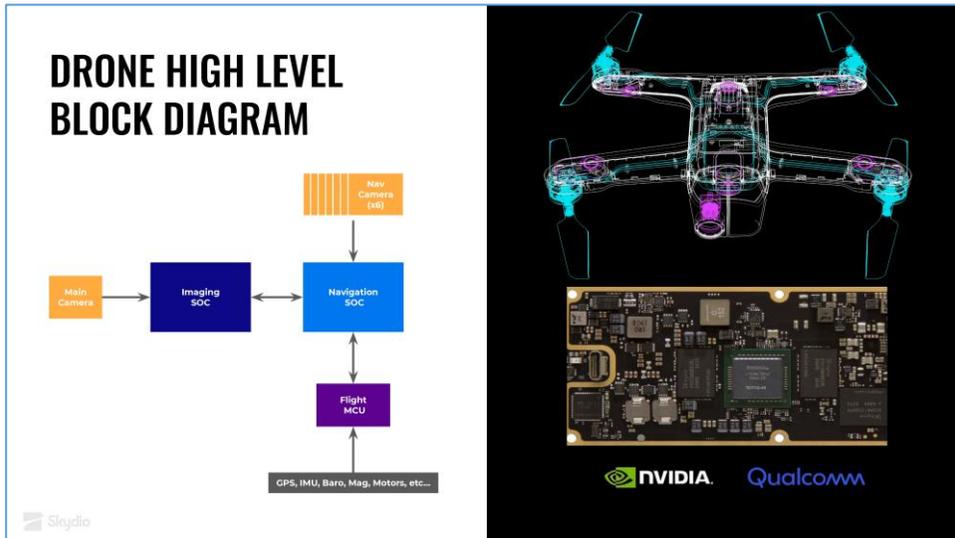
4-1-2. K2: Skydio Autonomy Engine: Enabling the Next Generation of Autonomous Flight (Abraham Bachrach, CTO, Skydio)

(1)本講演の概要

- 自律飛行にはイメージ処理、幾何的な処理、Deep Learning、Non linear Optimizationなど様々な計算処理タスクが必要。
 - 現状、Deep Learningだけでは非効率・不安定であり、古典的な計算手法との組み合わせ・適材適所化が重要である。
- ※Skydio: 自律飛行ドローンの分野で米国を代表するMIT AIラボ発のスタートアップ企業

(2)本講演が注目される理由

- 自律飛行に必要な計算タスクの分類と中身の説明は、チップ設計者にとって非常に参考になる内容である。
- 基本的には自動車における自動運転と似た処理内容であるが、自律飛行では大きさ・重量・消費電力の制約がより大きく、AI処理系の開発ドライバーターゲットとしてより先鋭的で優れている面があると思われる。
- あと一桁の性能/消費電力が向上すれば現状でやりたいことはできるとの報告だが、この向上は継続的に期待されると思われる。



4-2. 特筆すべきセッション論文

AI時代に合わせて従来の汎用機アーキテクチャに多くの工夫を施した発表、RISC-Vアーキテクチャの工夫でセキュリティを作り込むチップの安全性を向上させる発表、チャレンジング・主流とは違う独自の取り組みの発表、注目を集めていた製品の初めての技術詳細発表の下記5件をピックアップし、次スライド以降で発表のポイントについて解説する。

■ CPUs

論文番号	Title	Presenters
S1.3	The >5GHz next generation IBM Z processor chip	Christian Jacobi, IBM

■ Academic Spinout Chips

論文番号	Title	Presenters
S2.2	Morpheus II: A RISC-V Security Extension for Protecting Vulnerable Software and Hardware	Todd Austin, University of Michigan

■ Enabling Technologies

論文番号	Title	Presenters
S4.3	Aquabolt-XL: Samsung HBM2-PIM with in-memory processing for machine learning accelerators	Jin Hyun Kim, Samsung Electronics

■ ML Inference for the Cloud

論文番号	Title	Presenters
S5.1	Accelerating ML Recommendation with over a Thousand RISC-V/Tensor Processors on Esperanto's ET-SoC-1 Chip	David Ditzel, Esperanto Technologies

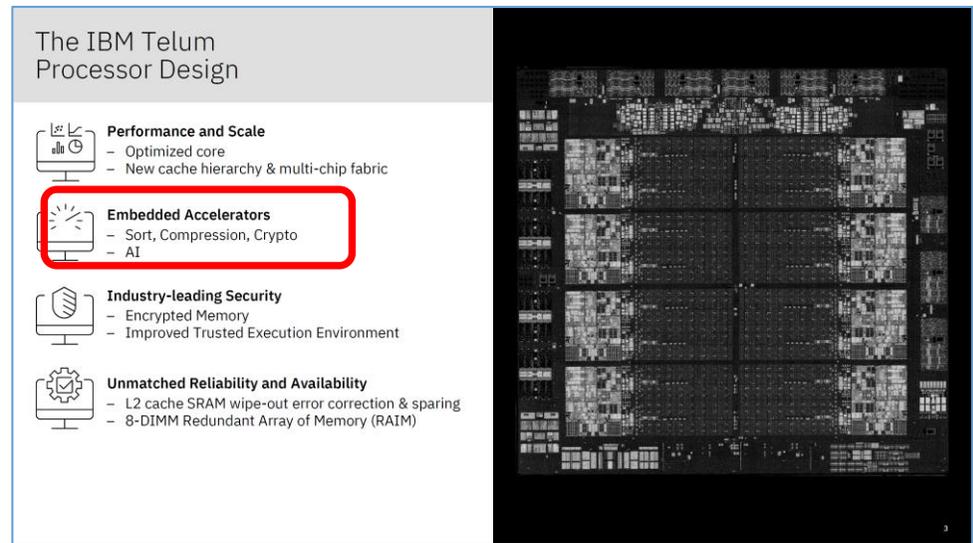
■ ML and Computation Platforms

論文番号	Title	Presenters
S6.1	Graphcore Colossus Mk2 IPU	Simon Knowles, Graphcore

4-2-1. S1.3: The >5GHz next generation IBM Z processor chip (Christian Jacobi, IBM)

(1)本論文の概要

- IBM Zメインフレームシリーズの最新世代チップ(Samsung 7nmプロセス)
 - ✓ メインフレーム市場でも既にAI応用が重要に(詐欺防止、顧客予測、サプライチェーン最適化、異常検知、等)
- コア間-チップ間のL2キャッシュを相互接続して**仮想的にキャッシュ容量を拡大(最大32Chips・8GBまで対応)**
- AIアクセラレータの搭載が大きな特徴
 - ✓ **メインフレーム内でAI処理を実行可能**
 - ✓ チップ内コアが一つのAIアクセラレータを共有
 - ✓ 従来型決定木、CNN、RNN等に対応
 - ✓ **128×の8-way FP16 SIMDからなる
シストリック型のMatrix Arrayと
32×の8-way FP16/32 SIMDからなる
Activation Arrayが特徴**
 - ✓ **ONNX(Open Neural Network Exchange)
を介してオープン環境に連結**



(2)本論文が注目される理由

- AIアクセラレータの構成自体は一般的ではあるが、**Shared-Cache型の8つのCPUコアでAIアクセラレータを共有してデータをやり取りする仕組みに工夫**がある。VLSI2021で発表済みのAIコアとは別のプロセッサ。

(3)本論文の技術・製品が応用される分野・製品およびニーズ等

- メインフレーム市場においてセキュリティおよび不正行為対策が重要になっており、本製品は**特に金融機関においてクレジットカードの不正利用やマネーロンダリングといった不正行為をトランザクション処理中にリアルタイムで判別する用途**が期待されている。

4-2-2. S2.2: Morpheus II: A RISC-V Security Extension for Protecting Vulnerable Software and Hardware (Todd Austin, University of Michigan)

(1)本論文の概要

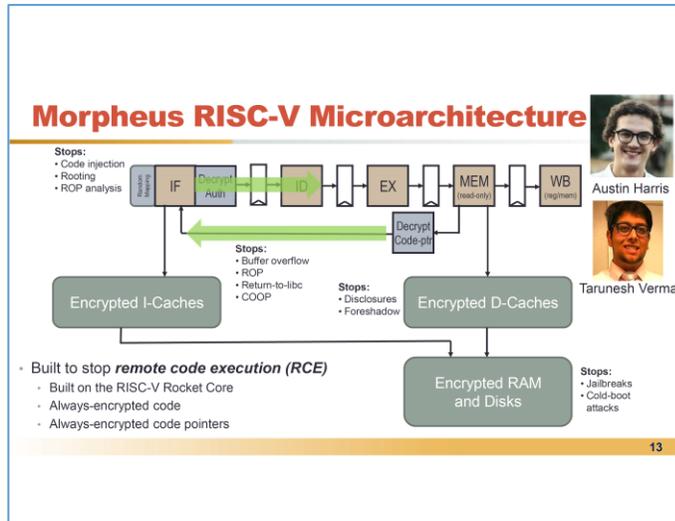
- プロセッサ内部実行側の機構に注目して攻撃に対抗するアプローチを提案
- RISC-Vの命令拡張、パイプライン構造の変更により、プログラム中の危うい箇所を定期的(50ms毎)にランダム化・暗号化する仕組みを導入
- 16nmプロセスでチップを試作し、性能・電力・面積オーバーヘッドは1%前後

(2)本論文が注目される理由

- 命令拡張がしやすいRISC-Vの特徴を有効活用している有望なアプローチである。
ただ、プログラム開発効率に対する懸念はある。
- 多面的な攻撃に対しては多面的な対策が必要であり、その中の一つの有効なアプローチである。

(3)本論文の技術・製品が応用される分野・製品およびニーズ等

- プロセッサ内のデータを命令およびキャッシュを含めて50msという短時間で異なるキーで再暗号化し、攻撃者に攻撃の手がかりを渡さないプロセッサであるため、サイバー攻撃に対して敏感な業界のニーズを満たすことができる。



Morpheus Design Overheads

- Integrated into the RISC-V Rocket Core
 - Only 369 lines of Chisel code added
- Deployed in a Xilinx UltraScale+ FPGA
 - Utilized a 12-round Simon cipher
 - < 1% performance overhead
 - 0.2% power overhead
 - 1.3% area overhead
 - Negligible impact to network apps



15

4-2-3. S4.3: Aquabolt-XL: Samsung HBM2-PIM with in-memory processing for machine learning accelerators (Jin Hyun Kim, Samsung Electronics)

(1) 本論文の概要

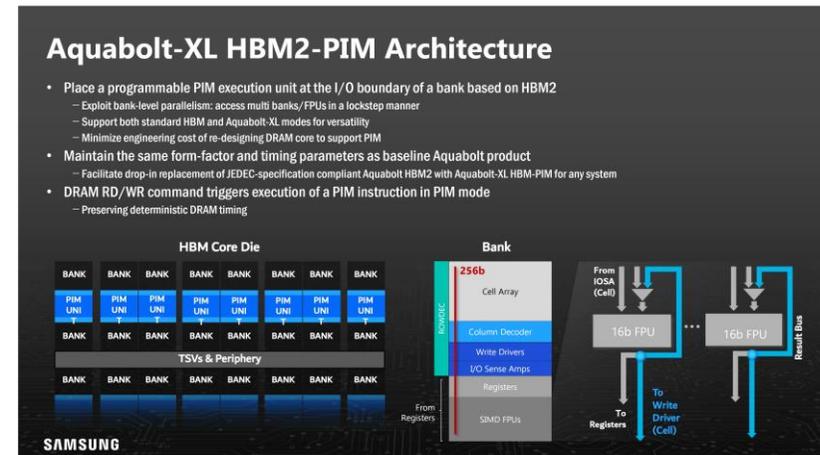
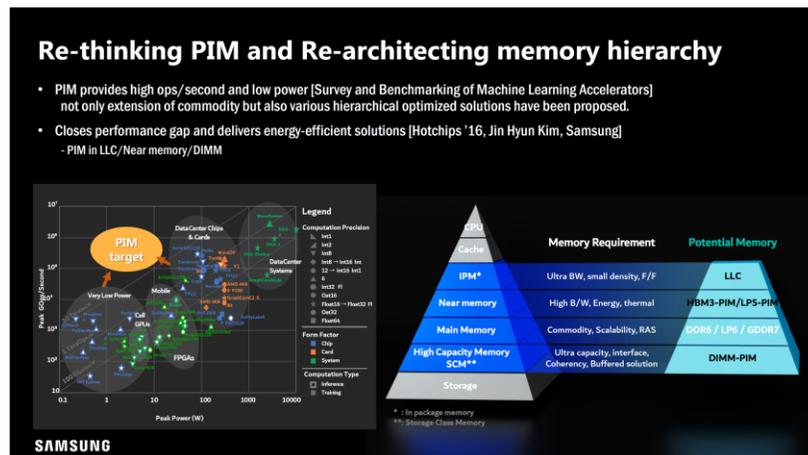
- メモリボトルネックになりがちで並列性も高いDNN処理が主ターゲットとするPIM(Processing in Memory)
- デモンストレータとして**3次元積層DRAMのHBM2の活用**にチャレンジ
- FP16の演算器をメモリバンクの出口に配置して並列計算実行

(2) 本論文が注目される理由

- Samsungがここ数年研究に力を入れている**メモリ移動の消費電力を低減するPIM(Processing in Memory)の最新研究成果の発表**である。
- DRAMとCPUが密に実装されるようになり、**AI処理が重要になった現在ならではのアプローチ**
- ロジックとメモリの付加価値の取り合いに勝てるか(どちらが主役になるのか)という観点での競争は注目される
残る課題は「DRAMは安い」という前提を覆す経済性の克服

(3) 本論文の技術・製品が応用される分野・製品およびニーズ等

- 処理するデータ量が増大しているAI処理において、メモリ帯域を増加させるニーズがある。また、同時に低消費電力化のニーズも大きい。そのため、本技術はHBM2をPIMに活用することで**メモリ帯域と低消費電力化のニーズに適うもの**となっている。**AI処理が必要な広い分野での利用**が考えられる。



4-2-4. S5.1: Accelerating ML Recommendation with over a Thousand RISC-V/Tensor Processors on Esperanto's ET-SoC-1 Chip (David Ditzel, Esperanto Technologies)

(1) 本論文の概要

- 並列RISC-VコアベースのAIアクセラレータであり、マルチスレッドでより多くのピンをメモリ/Oに使用してDRAMアクセスレイテンシーを隠蔽する従来手法を採用
- 10mWでRISC-Vコアが動作するように回路からアーキテクチャまで総合的に最適化、0.4V動作で効率向上。Tensor命令拡張を実施
- RISC-Vチップとして最高性能を謳う。ET-SoC-1チップに1093個のコアを搭載。データセンタ向け商用チップ ※7nmプロセス

(2) 本論文が注目される理由

- アーキテクチャ業界のレジェンドの一人であるDitzel氏の発表であり、AI系の主流アプローチである演算器をアレイ化するアプローチと違うマルチコアRISC-Vで高エネルギー効率を達成している。
- 今後、大規模演算器アレイ型のDNN処理との比較情報等を注視すべきチップである。

(3) 本論文の技術・製品が応用される分野・製品およびニーズ等

- AIの応用分野の一つである広告・リコメンド分野では推論能力に高いニーズがあり、10ms以下での応答が求められる。この分野をターゲットに開発された本製品は省エネ設計により省電力化も実現しており他の分野でも活用が見込まれる。

Esperanto's approach is different... and we think better for ML Recommendation

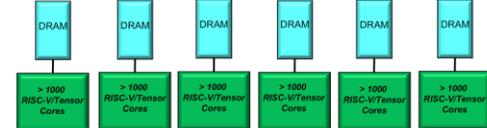
Other ML Chip approaches



10-20 CPU cores
and
Systolic Array Multipliers

DRAM

Esperanto's better approach



DRAM

> 1000 RISC-V/Tensor Cores

Use **multiple low-power** chips that still fit within power budget
Performance, pins, memory, bandwidth **scale up with more chips**

Thousands of general-purpose RISC-V/tensor cores

- Far more programmable than overly-specialized (eg systolic) hw
- **Thousands of threads** help with large sparse memory latency
- **Full parallelism** of thousands of cores always available
- **Low-voltage** operation of transistors **is more energy-efficient**
- Lower voltage operation also reduces power
- Requires both **circuit and architecture innovations**

Challenge: How to put the highest ML Recommendation performance onto a single accelerator card with a 120-watt limit?

Summary Statistics of ET-SoC-1

The ET-SoC-1 is fabricated in TSMC 7nm

- 24 billion transistors
- Die-area: 570 mm²
- 89 Mask Layers

1088 ET-Minion energy-efficient 64-bit RISC-V processors

- Each with an attached vector/tensor unit
- Typical operation 500 MHz to 1.5 GHz expected

4 ET-Maxion 64-bit high-performance RISC-V out-of-order processors

- Typical operation 500 MHz to 2 GHz expected

1 RISC-V service processor

Over 160 million bytes of on-die SRAM used for caches and scratchpad memory

Root of trust for secure boot

Power typically < 20 watts, can be adjusted for 10 to 60+ watts under SW control

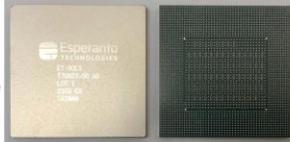
Package: 45x45mm with 2494 balls to PCB, over 30,000 bumps to die

- Each Minion Shire has independent low voltage power supply inputs that can be finely adjusted to mitigate V_t variation effects and enable DVFS

Status: Silicon currently undergoing bring-up and characterization



ET-SoC-1 Die Plot



ET-SoC-1 Package

4-2-5. S6.1: Graphcore Colossus Mk2 IPU (Simon Knowles, Graphcore)

(1) 本論文の概要

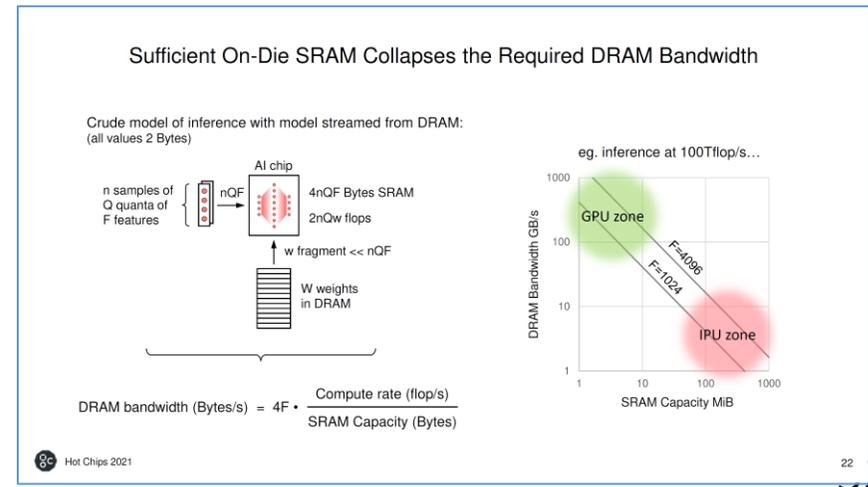
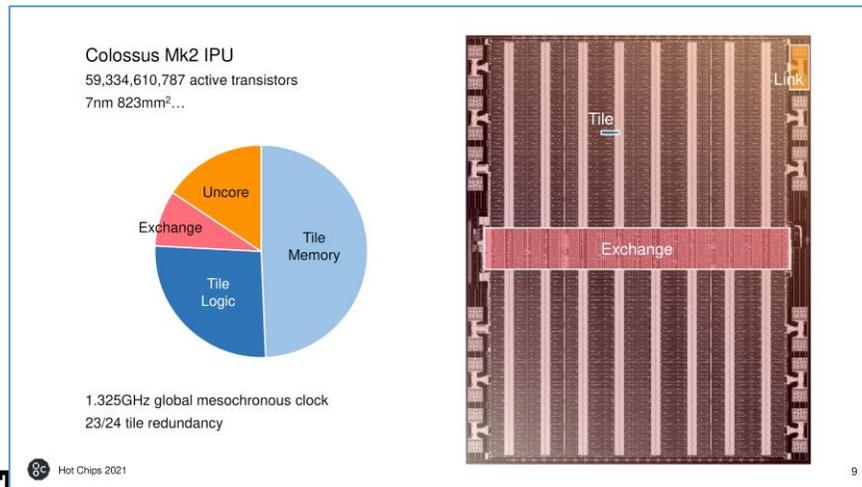
- データフロー型の実行モデルを採用した第2世代のGraphcore IPU(TSMC 7nmプロセス、1.3GHz)の発表
- 巨大なチップサイズ(2cm×4cm程度)の半分を占める**896MBの巨大なオンチップメモリ(47TB/s)**
- SupervisorとWorkerスレッドx6が存在。前者がパイプラインに常駐、後者はインタリーブすることで実効的に1/6の周波数で動作(プロセッサとしてはかなり特殊な構成)
- 分散SRAMを中心とするアーキテクチャによりエネルギー効率向上

(2) 本論文が注目される理由

- これまであまり具体的内容が明かされなかったGraphcoreアーキテクチャの思想・構成が示された発表
- 高コストのHBMは避け、サーバCPU越しのDRAMを利用するAIサブシステムとしてはユニークなアプローチ
- 全体的に非常に凝った賢いハードウェアアーキテクチャではあるが、GPU向けの最適化とは全く別の最適化が求められ、これにエコシステムがついてこられるか、今後注目される。
- 第1世代からの大きな方針転換。オンチップメモリ規模拡大、部分結合・非同期型への転換

(3) 本論文の技術・製品が応用される分野・製品およびニーズ等

- 処理するデータ量が増大しているAI処理、特にトレーニングにおいて、処理性能の向上、同時に低消費電力化のニーズは大きい。競合製品であるNVIDIAのGPUとのベンチマークを示して優位性をアピールしている。第1世代のIPUはMicrosoftのAzureで採用されており、**広範囲なAI応用の加速に使われる。**



4-3. 特筆すべきポスター論文

日本のAIチップに関する研究開発の現状を示す、日本に関連する発表（日本に拠点を置く珍しいAIチップスタートアップの発表、日本の大学として唯一AIチップ分野で成果を出しているグループの発表）2件を特筆すべきポスター論文としてピックアップした。

論文番号	Title	Authors & Affiliation
P04	Dynamic Neural Accelerator for Reconfigurable and Energy-efficient Neural Network Inference	Sakyasingha Dasgupta; EdgeCortix
P13	Edge Inference Engine for Deep & Random Sparse Neural Networks with 4-bit Cartesian-Product MAC Array and Pipelined Activation Aligner	Kota Ando; Tokyo Institute of Technology

4-3-1. P04: Dynamic Neural Accelerator for Reconfigurable and Energy-efficient Neural Network Inference (Sakyasingha Dasgupta; EdgeCortix)

(1)本論文の概要

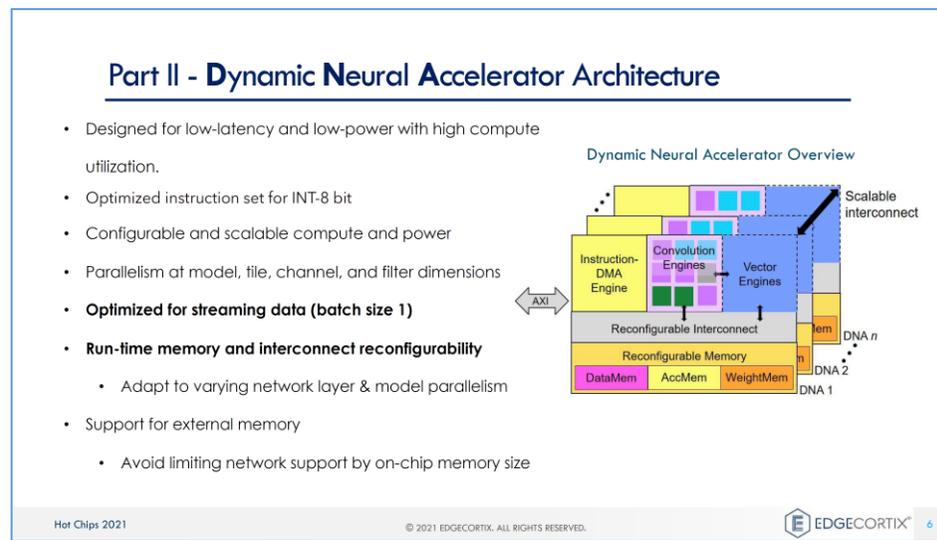
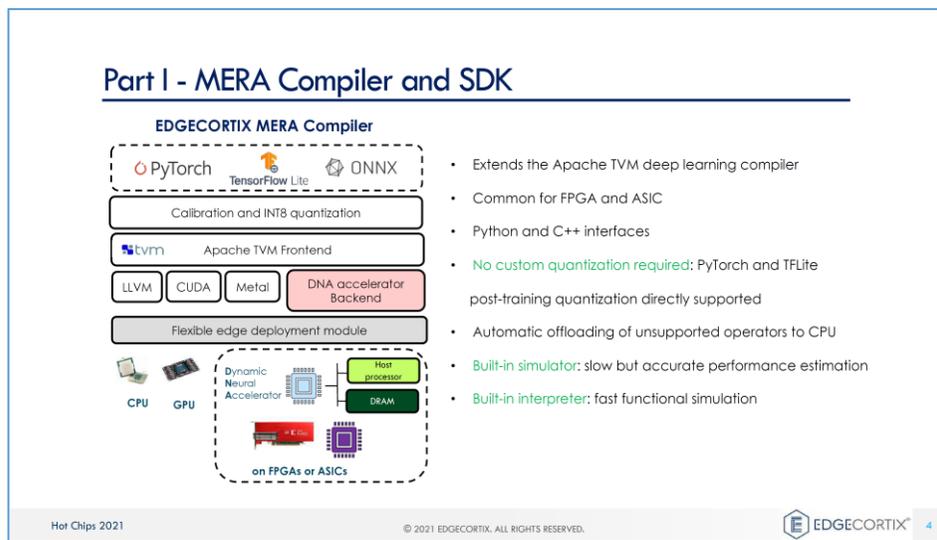
- DNN向け開発ソフトウェア(MERAコンパイラ)と実行ハードウェア(DNA)の両方をビジネスターゲットとした発表
- MERAコンパイラはPyTorchとTensorFlow Liteのトレーニング後量子化等を直接サポート
- DNAはコア-メモリ間の接続を回路レベルで再構成する仕組みであり、特徴となっている。

(2)本論文が注目される理由

- 米国系の人材が日本でスタートアップしている珍しいパターンであり、日本の市場特性や人材特性を強みに成長できるか注目される。
- 技術的にはハードウェアの動的再構成をポイントにしており、日本の技術土壌にフィットしている。

(3)本論文の技術・製品が応用される分野・製品およびニーズ等

- データセンタでのAI処理増加により消費電力の大きさが課題となっており、エッジでのAI処理が注目されている。本製品は低遅延・低消費電力に優れたFPGAに既存のフレームワーク(PyTorch, TensorFlow Lite)で開発されたDNNを比較的容易に配置することができ、エッジAIの性能向上のニーズがある分野に適合する。



4-3-2. P13: Edge Inference Engine for Deep & Random Sparse Neural Networks with 4-bit Cartesian-Product MAC Array and Pipelined Activation Aligner (Kota Ando; Tokyo Institute of Technology)

(1) 本論文の概要

- エッジ向け軽量DNNで多用されるDepth-Wise/Point-Wiseの積和演算に着目し、Depth-Wiseをシフト演算の重ね合わせで実現。更にDepth-Wiseフィルタのランダム枝刈りによりシフト演算の重ね合わせ数が減少し、実行効率が向上することを新規に提案し、チップとして実現
- 40nmチップにより26.5TOPS/W(枝刈り後)と非常に高いエネルギー効率を実現

(2) 本論文が注目される理由

- DNNアルゴリズムとハードウェアの協調研究が新しい。
- AIチップに関して米国・韓国から多数の産学発表がある中、日本からは産業界含めても本発表だけであり、日本の大学からこのような研究がさらに出てきて活性化することが期待される。

(3) 本論文の技術・製品が応用される分野・製品およびニーズ等

- アプリケーションの高度化に伴い大規模化したDNNは応用範囲が広がる一方、エッジ端末で利用する場合の計算量とメモリ容量の増大に対応するニーズが多くなった。本技術の適用により試作したチップで効果を確認しており、スマートフォンやIoT端末に適合する。

Shift-based convolution

Any Conv can be represented as superposition of "Shift-PW"

Depth-wise Conv is (almost always) followed by point-wise Conv

DW-PW can be represented as $\sum(PW(\text{Shift}(\cdot)))$

Hot Chips 33 8

Prototype chip

Fabricated a prototype chip

- 32x32 4-bit MAC array
- 534 MHz, 3.03 TOPS/W (dense) at 1.1V
- 234 MHz, 5.30 TOPS/W (dense) at 0.8V

Shift operations are not counted as OP:
Achieves much higher "sparse TOPS/W" with highly sparse models
✓ 5.30 TOPS/W (dense; measured)
= 26.5 TOPS/W (sparse; effective) @0.8V assuming the model is pruned to 1/9

Technology	TSMC 40nm CMOS (LP)
Chip size	3mm x 3mm
Core area	SRAM: 3.6mm ² Logic: 0.9mm ²
Core V _{DD}	0.75 - 1.2V
Frequency	534MHz@1.1V 234MHz@0.8V
Power Consumption	352mW@1.1V 88mW@0.8V
Gate Count	746K Gate
SRAM	AMem: 16k x 32 x 4 WMem: 16k x 32 x 4b IMem: 16k x 64b -- Total 5Mb

Hot Chips 33 18

【参考1】チュートリアルセッション発表概要

開催初日の22日に開催された「ML Performance and Real World Applications」・「Advanced Packaging」の2つのチュートリアルセッションの個々の発表の概要を次スライド以降に示す。

参考1-1. ML Performance and Real World Applications(1/4)

#	Title	Presenters
1	Modern Neural Networks and their Computational Characteristics	Paulius Micikevicius, NVIDIA
概要		
<p>この講演では、コンピュータビジョン、言語、音声、レコメンデーションといった主要なアプリケーションドメインの特徴とトレンドを見ながら、最新のDNNワークロードの計算上の特徴を調査する。畳み込みバリエーション、アテンションモジュール、リカレントセル、埋め込み、正規化など、これらのネットワークで発生する操作の主なカテゴリを確認する。また、入力データの性質(通常のグリッド、シーケンス、グラフ、非構造化)がDNNモデルアーキテクチャと操作の選択にどのように影響するかを調べる。最後に、与えられたネットワークの負荷が、入力特性、操作融合、および量子化とスパース性のような負荷低減技術の変化により、その訓練と推論の間でどのように異なるかを概説する。</p>		
#	Title	Presenters
2	MLPerf™ Training and Inference	Peter Mattson, Google
概要		
<p>MLPerfトレーニングおよび推論ベンチマークは、機械学習システムのパフォーマンス(速度)を測定するための業界標準になった。本論文では、機械学習性能のベンチマークにおける設計選択と、MLPerfトレーニングおよび推論ベンチマークがそれらの選択をどのようにナビゲートするかについて述べる。我々は、潜在的な提出者が円滑な提出を可能にすることを目標に、ベンチマークの提出とレビューのプロセスを順を追って説明する。ベンチマークスイートの2年以上の実績に示されているように、業界の進捗状況をレビューする。最後に、ベンチマーク・スイートを改善するための進行中の作業と、新しいコラボレーターがどのように関与し、フィールド全体に影響を与えるかについて説明する。</p>		
#	Title	Presenters
3	Software/hardware co-optimization on the IPU: An MLPerf™ case study	Mario Michael Krell, Graphcore
概要		
<p>機械学習は完全なシステム問題であり、ソフトウェアとハードウェアを横断した注意深い最適化を必要とする。このケーススタディーでは、現実世界のAIモデルにおけるすべてのアーキテクチャおよびシステムエンジニアリングイノベーションの集大成である、MLPerf v 1.0トレーニングベンチマークへのGraphcoreサブミッションのパフォーマンス結果と主要な最適化を示す。NLP(自然言語処理)および画像分類における一連のモデルのための最適化された実装を提供する。これらのモデルをダイに適合させる能力の中心は、モデル並列性、FP 16マスタ重み、外部ストリーミング・メモリ、小バッチサイズ訓練を含む新しい技術である。</p>		

(資料) <https://hotchips.org/>

参考1-1. ML Performance and Real World Applications(2/4)

#	Title	Presenters
4	Deep Learning Inference Optimizations on CPUs	Guokai Ma, Intel
概要		
<p>ディープラーニング(DL)推論アプリケーションは、接続されたデバイスからのデータが大幅に増加するにつれて急速に成長している。専用のディープラーニングアクセラレータがあるが、CPUは今日でも最も利用可能な推論プラットフォームである。CPUの最適化は、レイテンシーの削減やスループットの向上により、ディープラーニングアプリケーションに直接的な時間とコストの節約をもたらす。ビジョンモデルとは異なり、ほとんどの言語、音声、および推奨モデルは、広範囲の入力形状を受け入れ、サイズが非常に大きくなる可能性がある。これらのユースケースを最適化するために、次の5つの手法を使用した。(1)低精度の推論。(2)スパース性を導入することによって計算を削減。(3)opsフュージョンによるメモリアクセスの削減。(4)プリミティブ作成のオーバーヘッドを削減。(5)入力サイズの負荷分散とより多くの並列処理を導入することにより、ハードウェア使用率を改善。いくつかの特定のDLモデルを検討するが、それぞれがDLモデルのより一般的なクラスの例であり、これらの最適化が適用されることが期待される。最後に、すべての実装の詳細は、インテルの最新のMLPerf™推論v1.0の提出物でオープンソース化されている。</p>		

#	Title	Presenters
5	AI at Scale for the Modern Era	Carole-Jean Wu & Niket Agarwal, Facebook
概要		
<p>過去10年間で、AIの計算量は30万倍に増加した。最新の自然言語処理モデルには1兆を超えるパラメータが含まれている一方で、ニューラル・レコメンデーションおよびランキング・モデルのメモリー・ニーズは数百ギガバイトからテラバイト規模に拡大している。最先端の業界規模のパーソナライゼーションとレコメンデーションモデルのトレーニングは、Facebookのすべてのディープラーニングのユースケースの中で最も多くの計算サイクルを消費する。業界規模のディープラーニングモデルが直面するシステム上の主な課題は何か?この講演では、MLデータの前処理からトレーニングシステムのスループット最適化に至るまで、機械学習実行パイプラインの端から端まで、インフラストラクチャ最適化の課題と機会に対する規模と影響に焦点を当てる。講演は、高性能で効率的なAIシステムを大規模に構築するための方向性で締めくくられる。</p>		

(資料) <https://hotchips.org/>

参考1-1. ML Performance and Real World Applications(3/4)

#	Title	Presenters
6	The Nature of Graph Neural Network Workloads	Da Zheng and George Karypis, Amazon
概要		
<p>グラフニューラルネットワーク(GNN)はグラフデータから学習するための強力なツールであり、不正検出やレコメンデーションなどの様々なアプリケーションで広く使用されている。産業におけるGNN研究と採用の人気は、GNN訓練を最適化するための多くのシステム研究に火をつけた。GNN計算の特性は3つの要因に依存する: (i) GNNモデルのアーキテクチャ、(ii) グラフトポロジとそれに関連するノード/エッジ特徴のタイプ、および (iii) 訓練アルゴリズム。初期のGNNシステム研究は、GraphSageやGCNのような単純なGNNアーキテクチャを、フルバッチトレーニングを伴う小さな同質グラフ上で最適化することに焦点を当てた。しかし、GNNワークロードは、複雑なアーキテクチャを持つGNNモデル、不均一で多様な情報を持つグラフ、およびミニバッチトレーニングに向かって移動している。新しいワークロードに向けたGNNフレームワーク最適化とハードウェア設計を導くために、単純および複雑なGNNアーキテクチャ、異なるタイプおよび異なるスケールのグラフデータセット、および様々なトレーニング方法の両方をカバーするGNNベンチマークを開発している。本稿では、GNNベンチマークの最初のバージョンと、各種ハードウェアプラットフォーム上のGNNワークロードの特性を示す。</p>		

#	Title	Presenters
7	Challenges in large scale training of Giant models on large TPU machines	Sameer Kumar, Google
概要		
<p>この講演では、巨大なTPUマシン上で巨大なモデルをスケールリングし、トレーニングする上でのさまざまな課題について説明する。最初に、MeshTensorFlow、GShardなどの大規模な言語モデルを並列化するための大規模な手法をレビューする。また、モデル開発、パフォーマンス最適化、マシンの可用性、ネットワークポロジ、デバッグの課題、および巨大なモデルを提供するための手法についても説明する。デバッグの大きな課題は、同じモデルをより小規模に実行できるようにすることであり、パフォーマンスの最適化手法には、通信と計算をオーバーラップさせる積極的な手法による通信オーバーヘッドの最適化が含まれる。</p>		

(資料) <https://hotchips.org/>

参考1-1. ML Performance and Real World Applications(4/4)

#	Title	Presenters
8	ZeRO-Infinity and DeepSpeed: Breaking the device Memory Wall for Extreme Scale Deep Learning	Yuxiong He and Samyam Rajbhandari, Microsoft
概要		
<p>ZeRO-Infinityは、ディープラーニングのモデルトレーニングをスケールアップするための新しいディープラーニング (DL) トレーニング技術で、単一のGPUから数千のGPUを備えた巨大なスーパーコンピューターまで対応する。システムの全メモリ容量を活用し、モデルコードのリファクタリングを必要とせずすべての異種メモリ(GPU、CPU、およびNon-Volatile Memory Express (略してNVMe))を同時に活用することで、前例のないモデルサイズを強化する。同時に、制限されたCPUやNVMeの帯域幅に煩わされることなく、優れたトレーニング・スループットと拡張性を実現する。ZeRO-Infinityは、現行世代のGPUクラスタのトレーニング用に、数十兆から数千兆のパラメータを持つモデルにフィットさせることができる。単一のNVIDIA DGX-2ノードで1兆個のパラメータモデルを微調整できるため、大規模なモデルへのアクセスが容易になる。トレーニングスループットとスケーラビリティに関しては、512個のNVIDIA V 100 GPU (ピーク時の40%) で25ペタフロップスを維持し、超線形スケーラビリティも実証している。ZeRO-Infinityのオープンソース実装は、DeepSpeed.aiから入手できる。</p>		

(資料) <https://hotchips.org/>

参考1-2. Advanced Packaging(1/2)

#	Title	Presenters
1	Technology Provider: Intel packaging technologies for chiplets and 3D	Ravi Mahajan and Sandeep Sane, Intel
	概要	
	<p>先進的なパッケージング・テクノロジーは、コンパクトで電力効率に優れたプラットフォームとして重要であるため、異機種混在環境の統合 (HI) を実現するうえで不可欠である。この講演では、HIプラットフォームとしてのパッケージングの価値を確立し、さまざまなパッケージングアーキテクチャの機能について説明する。これらのアーキテクチャは、主に物理的な相互接続機能に基づいて比較される。EMIB、Silicon Interposer、Foveros、Co-EMIBなど、最先端の2Dおよび3Dテクノロジーの主要な機能について説明し、それらの進化のロードマップを示す。堅牢で高度なパッケージアーキテクチャーを開発する際の課題と機会については、インテルが最先端のパッケージングテクノロジーをグラフィックス、クライアント、およびFPGAアプリケーションで使用する場合との関連で説明する。</p>	
#	Title	Presenters
2	Technology Provider: TSMC packaging technologies for chiplets and 3D	Doug Yu, TSMC
	概要	
	<p>確立されたTSMC 3DFabric™テクノロジープラットフォームを使用して、異種システムパッケージエンベロープの拡張とシステム相互接続の縮小を続けている。このシステム相互接続スケーリングは、3D相互接続密度、帯域幅/待ち時間およびエネルギー効率性能 (EEP) を駆動するために提案したロードマップに基づいている。また、ウェーハレベルシステムインテグレーション技術を活用して、液冷システムへ移行する際の放熱性を高める革新的なソリューションを提供している。さらに、HPCのEEPとネットワークを駆動するためのSiフォトニクス応用のための破壊的コンパクトユニバーサルフォトニックエンジン (COUPE) を、システムEEPを駆動するために新たに導入した。結果はここで共有される。</p>	

(資料) <https://hotchips.org/>

参考1-2. Advanced Packaging(2/2)

#	Title	Presenters
3	Case Study: Intel products built with 2.5D and 3D packaging	Ravi Mahajan and Sandeep Sane, Intel

#	Title	Presenters
4	Case Study: AMD products built with 3D packaging	Raja Swaminathan, AMD

概要

チップレットアーキテクチャが主流となり、省電力コンピューティングの継続的で経済的な成長を可能にするための基礎として認識されるようになるにつれ、高度なパッケージング技術とアーキテクチャは、異種統合によってムーアの法則の次のフロンティアを可能にするために、より重要になってきている。このチュートリアルでは、AMDがPPAC (電力、性能、面積、コスト)の改善を可能にし、異機種混在アーキテクチャを可能にする高度なパッケージアーキテクチャーについて説明する。AMDの3D VCacheアーキテクチャで使用されているダイレクトCu-Cuボンディング技術は、詳細に説明されており、PPACの利点について業界標準の3Dアーキテクチャと比較される。ハイパフォーマンスコンピューティングアーキテクチャを発展させるために有効になっているその他のテクノロジーもプレビューされる。

#	Title	Presenters
5	Expert Opinion: An overview of the package technology landscape and industry deployment	Jan Vardaman, TechSearch International Inc

概要

異種集積、特にチップレットの役割は、エレクトロニクス実装のこの新しい時代において重要である。高パフォーマンス・パッケージには多くの選択肢がある。このプレゼンテーションでは、さまざまなアプリケーションに適用されるオプションについて説明し、それぞれのメリットを示す。

【参考2】 その他注目セッション論文

その他注目セッション論文として以下の8件を参考にピックアップした。それぞれのポイントを次スライド以降に示す。

■ CPUs

論文番号	Title	Presenters
S1.1	Intel Alder Lake CPU Architectures	Efraim Rotem, Intel
S1.2	AMD Next Generation "Zen 3" Core	Mark Evers, AMD
S1.4	Next-Gen Intel Xeon CPU - Sapphire Rapids	Arijit Biswas and Sailesh Kottapalli, Intel

■ Academic Spinout Chips

論文番号	Title	Presenters
S2.1	Mozart: Designing for Software Maturity and the Next Paradigm for Chip Architectures	Karu Sankaralingam, University of Wisconsin- Madison

■ Infrastructure and Data Processors

論文番号	Title	Presenters
S3.1	Arm Neoverse N2: Arm's second-generation high performance infrastructure CPUs and system products	Andrea Pellegrini, ARM
S3.2	NVIDIA DATA Center Processing Unit (DPU) Architecture	Idan Burstein, NVIDIA

■ ML Inference for the Cloud

論文番号	Title	Presenters
S5.3	Qualcomm Cloud AI 100: 12 TOPs/W Scalable, High Performance and Low Latency Deep Learning Inference Accelerator	Karam Chatha, Qualcomm Inc

■ ML and Computation Platforms

論文番号	Title	Presenters
S6.2	The Multi-Million Core, Multi-Wafer AI Cluster	Sean Lie, Cerebras Systems

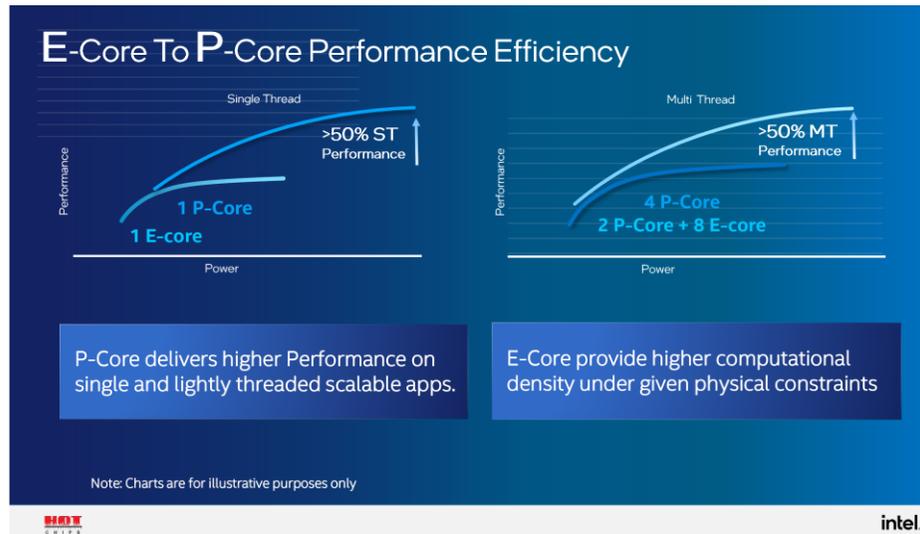
参考2-1. S1.1: Intel Alder Lake CPU Architectures (Efraim Rotem, Intel)

(1) 本論文の概要

- 高性能(P)コアと高効率(E)コアの組み合わせでチップを構成。X86アーキテクチャの中で最大級の進化を謳う。
- スマホからサーバーまでスケールラブルファミリーを実現している。
- 新技術スレッドダイレクター
 - ✓ OSレベルのスレッドスケジューラと連携することできめ細かいPコア/Eコア制御・タスク割当を実現
 - ✓ 目的は電力制御・エネルギー効率向上

(2) 本論文が注目される理由

- Arm社が2011年に提案したbig.LITTLEコンセプトのIntel版であり、スマホ向けSoCを中心に進んできた**エネルギー効率重視のチップの実現形態がIntel基幹CPUに波及してきた**と見ることができる発表である。
- スレッドダイレクターの中身はあまり開示されていないため新規性の判定は難しいが、ソフトとハードの連携で効率向上を目指す順当なアプローチをとっていると見ることができる。



Introducing Intel Thread Director
Intelligence built directly into the core

OS Scheduler

- Monitors the runtime instruction mix of each thread and as well as the state of each core – with nanosecond precision
- Provides runtime feedback to the OS to make the optimal scheduling decision for any workload or workflow
- Dynamically adapts guidance based on the thermal design point, operating conditions, and power settings – without any user input
- Power and energy management: Adjust Voltage and frequency to meet user experience while optimizing power, thermal and energy consumption

Alder Lake SoC

intel 14

参考2-2. S1.2: AMD Next Generation “Zen 3” Core (Mark Evers, AMD)

(1)本論文の概要

- ZENシリーズ3代目(ZEN2と同じ7nmプロセス)の発表
- 10命令同時発行。IPC(クロック当たりの命令数)で**ZEN2より19%向上**
- CCD(Core Complex Die)あたり8つのCPUコアと32MBのL3キャッシュを搭載
 - ✓ L3キャッシュはL2キャッシュと排他的(L2キャッシュから溢れたキャッシュをL3に保持)
- 更に、64MBのV-Cacheの三次元集積が可能
 - ✓ **TSV(Through-Silicon Via: シリコン貫通電極)による三次元積層**

(2)本論文が注目される理由

- コア内の工夫がよく分かる発表。
- 様々なアーキテクチャの工夫で**同一プロセスで約2割の実行効率向上と4.7->4.9GHzの周波数の向上を達成している。**
- **SRAMの3次元集積(V-Cache)によりサーバ/デスクトップクラスCPUの性能も大幅に向上できることを実証している。**

“ZEN 3” OBJECTIVES

PERFORMANCE

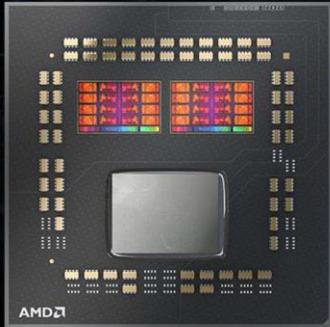
- Deliver another landmark increase in 1T performance through IPC and frequency
- Unify cores and cache in a contiguous 8-core complex to improve effective latency
- Provide scale-out performance for servers, data-centers and super-computers

NEW CAPABILITIES

- Introduce new ISA extensions
- Expanded security features
- Support for AMD 3D V-Cache integration

PLATFORM

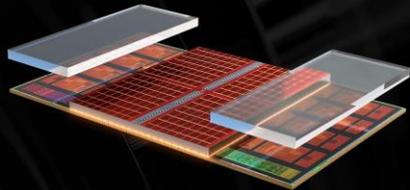
- Support for scaling and energy efficiency
- Socket compatibility



AMD

AMD 3D V-CACHE 192M L3 PROTOTYPE

- Zen 3 base CCD design includes 32M L3 cache
- Increased to 96M per CCD with 64M AMD 3D V-Cache
- Enabled by Through Silicon Vias on CCD
- Direct copper-to-copper bond



15% FASTER GAMING
ON AVERAGE*



Game	Performance Improvement
DOTA 2 (Vulkan*)	UP TO 1.18X
Gears V (DirectX® 12)	UP TO 1.12X
Monster Hunter World (DirectX® 11)	UP TO 1.25X
League of Legends (DirectX® 11)	UP TO 1.04X
Fortnite (DirectX® 12)	UP TO 1.17X

■ Ryzen™ 9 5900X / 4.0 GHz Fixed Frequency ■ 12-Core 3D Chiplet Prototype / 4.0 GHz Fixed Frequency

* See endnote: R5K-078.

AMD

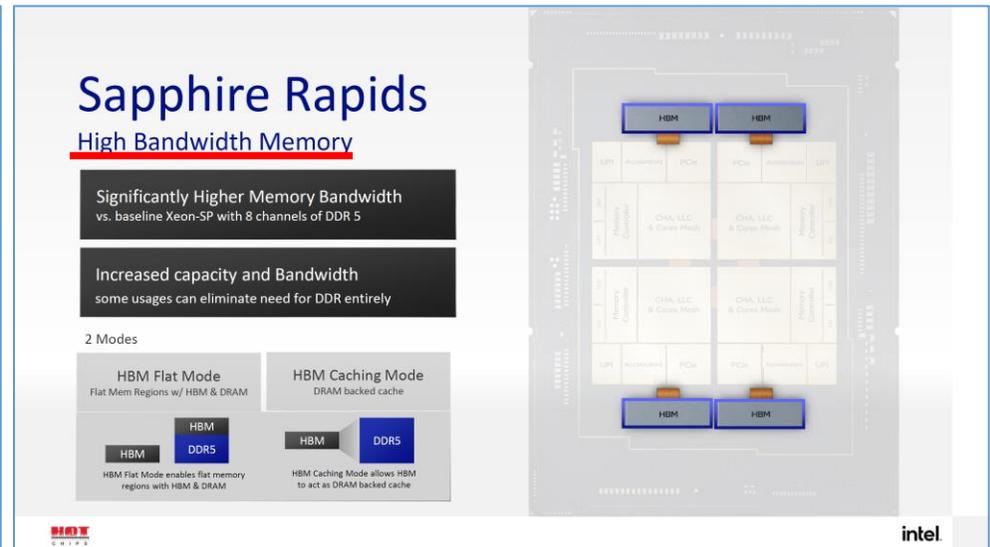
参考2-3. S1.4: Next-Gen Intel Xeon CPU - Sapphire Rapids (Arijit Biswas and Sailesh Kottapalli, Intel)

(1) 本論文の概要

- チップレットのパッケージ内集積、モジュラーアーキテクチャを採用したデータセンタ向けXeonシリーズの最新世代チップ
- 各種アクセラレータを搭載(データストリーミングアクセラレータエンジン、暗号エンジン、負荷分散、等)
- HBM(3次元積層DRAMモジュール)を接続可能
 - ✓ キャッシュとして使うか、メモリの一部として使うかを選択可能
- AI向け機能：低ビット精度の表現・命令をサポート

(2) 本論文が注目される理由

- HBMの使い方が注目されており、会議でもHBMの使い方に集中した。ただし、会議では詳細が明かされず、今後の発表も注目される。



参考2-4. S2.1: Mozart: Designing for Software Maturity and the Next Paradigm for Chip Architectures (Karu Sankaralingam, University of Wisconsin-Madison)

(1)本論文の概要

- ソフトウェアコンパチビリティの視点から見たDNN向けアーキテクチャフレームワークの提案
- プログラムの基本機能をデータ収集、同期、計算、制御に分解してハードウェア要素にマッピング
- 16nmでチップ試作

(2)本論文が注目される理由

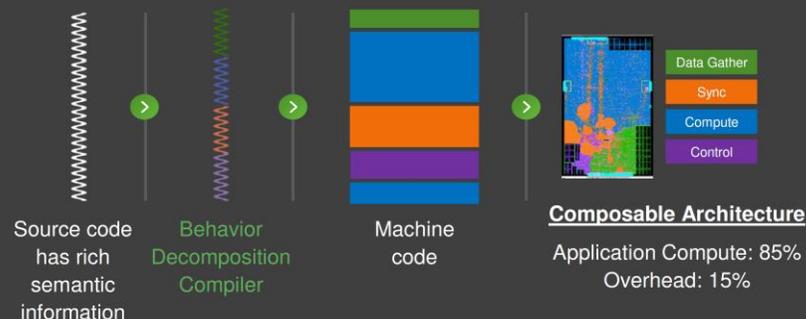
- GPUが非常に強力なAI研究プラットフォームであることを踏まえつつ、それだけでは問題が解決していないということ、新しいアーキテクチャにはチップができたその日から使えるzero-day開発環境が必要という問題提起は非常に正しい。
ただ、その解決策を定義しているレベルの抽象度が高すぎて、実問題への応用時に実問題の詳細にトラップされる懸念はある。

Is there a problem?

1. GPU chip utilization is quite low (10% of peak)
2. The software stack is turgidly tied to hand-tuned/auto-tuned libraries, resulting in ML systems getting caught in the rut.
3. Tail wagging the dog: Because only some styles of kernels achieve high performance on GPUs, ML practitioners are forced to create ridiculously large models to get new capability, rather than use better algorithms.
4. Simpler models exist but don't run fast on existing HW or SW
 - a. Lottery Hypothesis shows simpler models have the same/superior learning ability.
 - b. Depth-wise separable convolutions have 10X fewer ops, but provide no speedups on today's HW.

New chip architectures must have software maturity to challenge GPU dominance and serve as an alternative

Composable Computing Paradigm



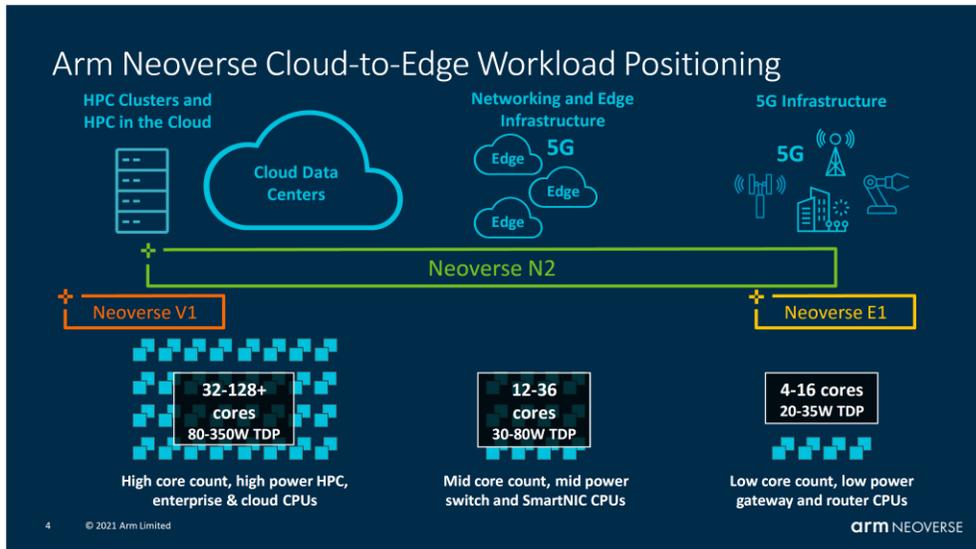
参考2-5. S3.1: Arm Neoverse N2: Arm's second-generation high performance infrastructure CPUs and system products (Andrea Pellegrini, ARM)

(1)本論文の概要

- Armv9アーキテクチャを初採用し、コア数調整によりクラウドからエッジまでカバーする戦略的IPコアの発表
- Marvell社のOCTEON DPU(Data Processing Unit)に採用されている。

(2)本論文が注目される理由

- Armの牙城である組み込みシステムにおいてローエンド側にRISC-Vが入ってくる動きがあり、Armにとってクラウド戦略の重要性はましており、本論文で発表のIPコアは**Armがクラウド側に進出するための戦略的IPコア**であるため。
- **Armをめぐる経済安保的側面、CPUコア技術の囲い込み戦略的観点も含め、このIPコアがどこまで浸透・伸長するか注目される。**



Arm Neoverse N2... Coming Soon in Silicon

Marvell OCTEON 10 Family	
Compute	<ul style="list-style-type: none">Up to 36 Neoverse N2 coresSPECint > 1000
Accelerators	<ul style="list-style-type: none">inline ML engineinline IPsecSSL/TLSVector packet processing (VPP)
SW	<ul style="list-style-type: none">DPDK networking suite for Control, Management and Fast path stacksSDK with Linux kernel and user plane extensions
Memory	<ul style="list-style-type: none">DDR5-5200 + ECC on-board memory
I/O	<ul style="list-style-type: none">Up to 400GEIntegrated 1Terabit switchPCIe 5.0

MARVELL
OCTEON 10
DPU Family

Available in Q4

© 2021 Marvell. All rights reserved.

© 2021 Arm Limited

arm NEOVERSE

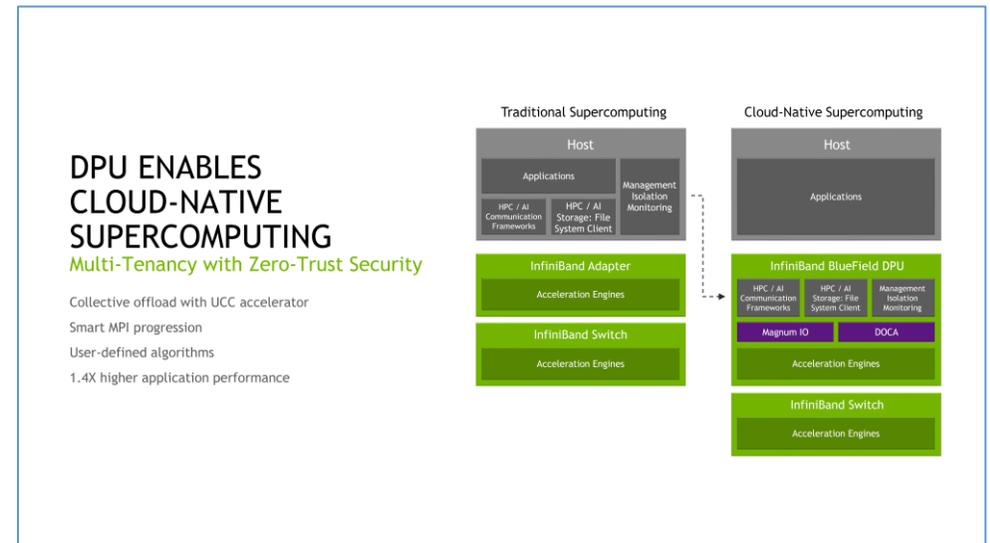
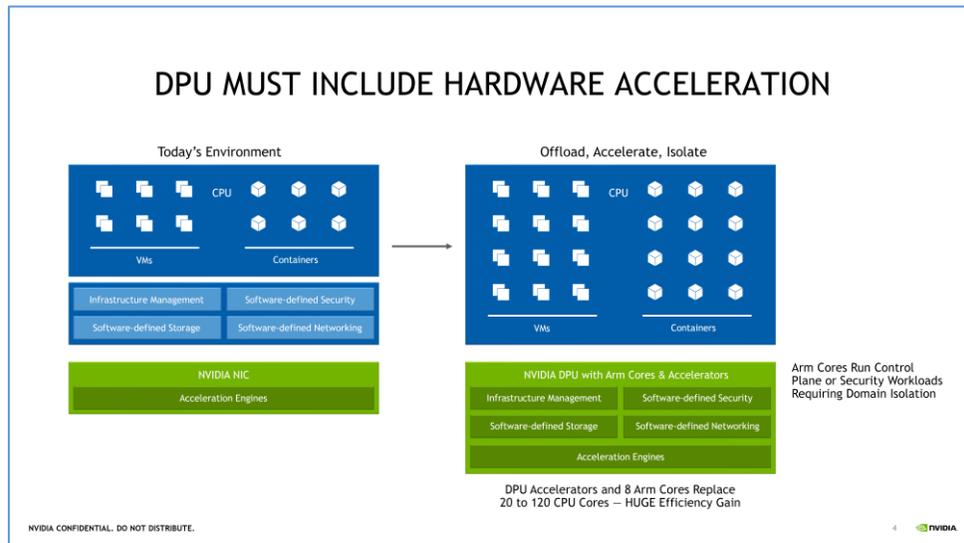
参考2-6. S3.2: NVIDIA DATA Center Processing Unit (DPU) Architecture (Idan Burstein, NVIDIA)

(1)本論文の概要

- GPUの次にNVIDIAがデータセンタへの採用を狙っているチップの発表
- ネットワーク近傍(NICないしはその拡張)でデータ処理を加速するコンセプト。Network Infrastructure Processingを担う。
- ARM+アクセラレータコンプレックス構成を採用

(2)本論文が注目される理由

- 発表後まだあまり時間がたっていないだけにGPUでの圧倒的な強みがどれだけ生きるか、今後の展開が注目される。
- NW側でパケット処理を行うことによるデータセンタ内でのSecurity Isolationの実効性がキーポイントとなる。



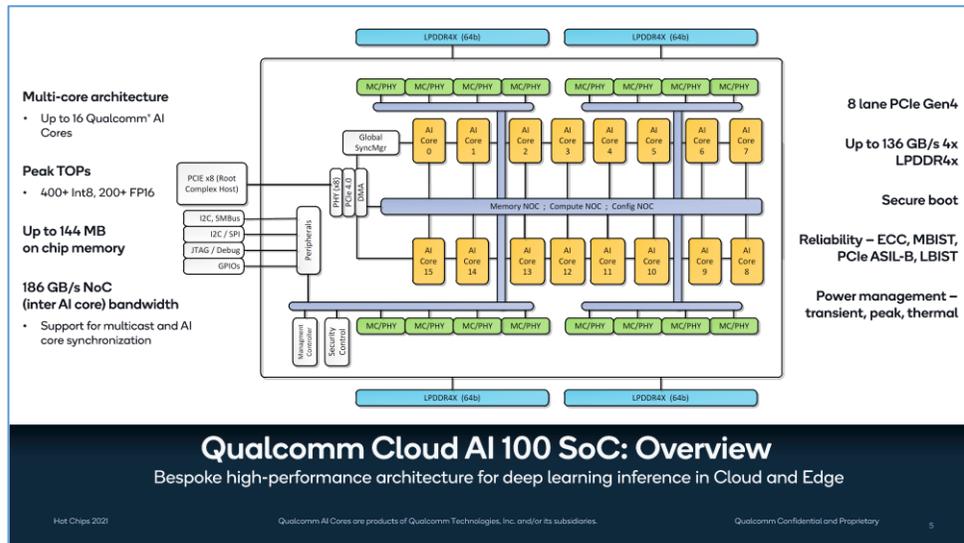
参考2-7. S5.3: Qualcomm Cloud AI 100: 12 TOPs/W Scalable, High Performance and Low Latency Deep Learning Inference Accelerator (Karam Chatha, Qualcomm Inc)

(1) 本論文の概要

- 12TOPS/Wを謳うクラウドとエッジ両方に向けたAI推論チップ
- VLIW型AIコア16個、8K個のMACユニット、NoC(Network on Chip)、144MBのオンチップメモリ
- ニューラルネットのレイヤをまたがって処理するレイヤフュージョンアプローチを採用し、コンパイラがこれに対応

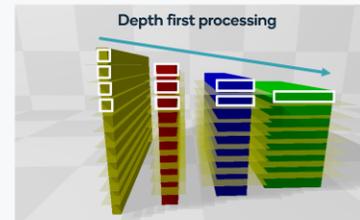
(2) 本論文が注目される理由

- Snapdragon向けに複数世代開発してきたAIコアのスケールアップ展開の位置づけのチップであり、既存のニューラルネットの推論の範囲内ではかなり熟成されたアーキテクチャである。
※TOPS/Wの値については、本論文での簡単な処理ではなくベンチマークの処理で見せるべきであった。



Low power optimizations

Compiler driven optimizations for improving power efficiency



Depth first scheduling

- Minimizes spillage to DDR by processing the network graph in a depth first manner
- Reduces DDR power consumption, and improves performance
- Particularly effective for large resolution images
- RN34-SSD - 5.3X lower DDR BW, 3.5X higher inf/sec, 2X higher inf/sec/W

Scheduling to minimize peak power excursions

- Schedules the network graph across ML cores such that computation is not correlated

Instruction usage-based power management

- Compiler generates metadata for power management controller

Hot Chips 2021

Qualcomm Confidential and Proprietary

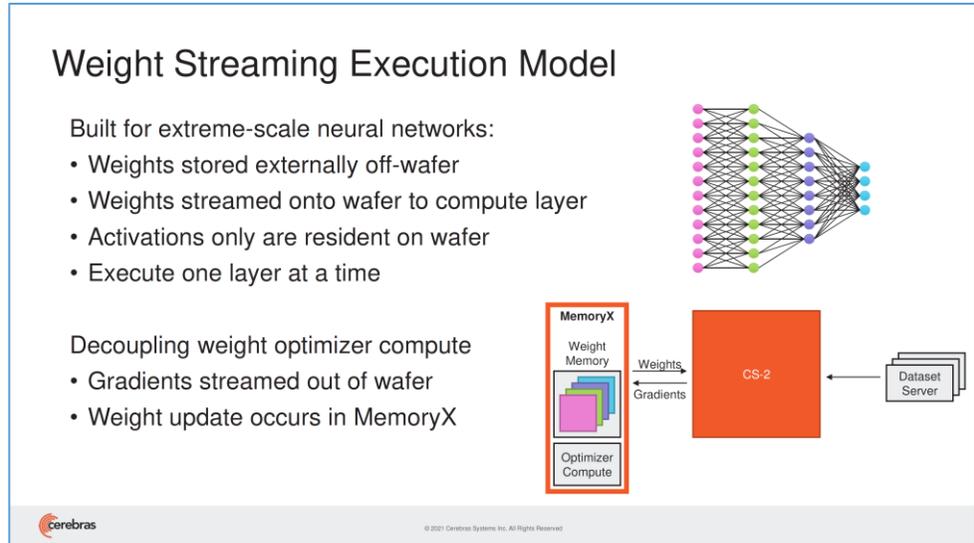
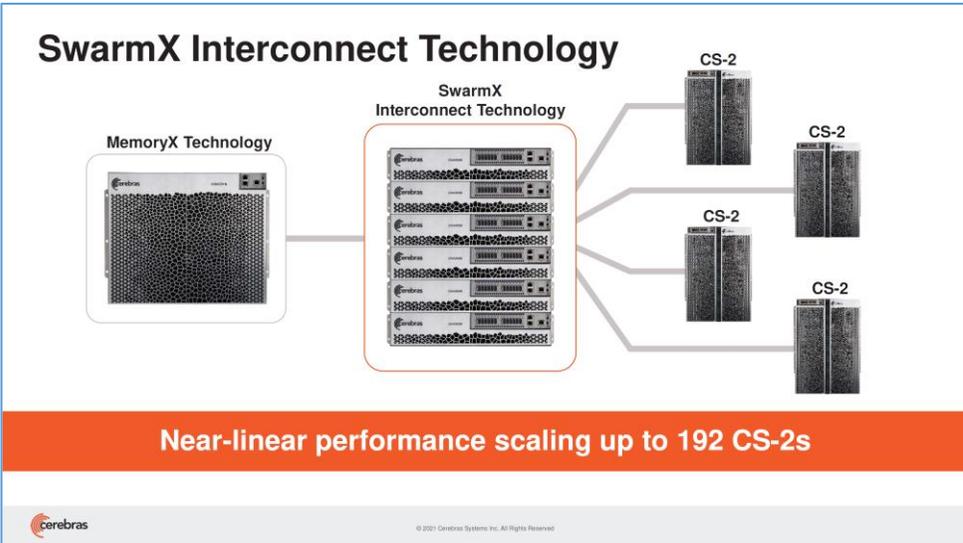
参考2-8. S6.2: The Multi-Million Core, Multi-Wafer AI Cluster (Sean Lie, Cerebras Systems)

(1)本論文の概要

- ウェーハスケールAIチップ(10万×10万規模の積和演算器アレイ)のクラスターの発表
- 推論時、重みを外部の専用メモリシステム装置(MemoryX)からストリームインし、計算結果をチップ内に保持。学習時はGradient(傾き)をストリームアウト
- MemoryXは4TB~2.4PBの容量を持ち、内部に重み更新等の計算機能を持つ。
- 重みのランダムスパースリティ対応も重視(スケーラブルに実現)

(2)本論文が注目される理由

- ウェーハスケールAIチップよりも、むしろ重み供給システムMemoryXの方に付加価値がシフトした発表
- 高コストマシンを欲する市場規模がどの程度あるか注目される。



【用語集】 (1/2)

スライド#	用語	意味
14	CAD, EDA	CAD(Computer-Aided Design)は、コンピュータ支援設計とも訳され、コンピュータを用いて設計をすること、あるいはコンピュータによる設計支援ツールのこと。 EDA(Electronic Design Automation)は、集積回路や電子機器など電気系の設計作業の自動化を支援するためのソフトウェアやハードウェアのこと。設計機能だけでなく、シミュレーションによる機能・特性の確認までできる。
14	ハイパーコンバージド統合	半導体設計の論理的設計と物理的設計を同一のモデルを使って行うことで、実装後の手戻りを減らす設計の考え方
17	SIMD	Single Instruction/Multiple Data(単一命令/多重データ)を略したもので、1つの命令を同時に複数のデータに適用する並列化の形態。ベクトル演算と呼ぶこともある。 並列化の形態として、この他にMIMD(Multiple Instruction stream/Multiple Data stream)もある。SIMDが単一命令であったのに対し、MIMDは命令も多重化されており、異なる命令・異なるデータが同時に処理される。
17	ONNX	Open Neural Network Exchangeの略。機械学習や人工知能のモデルを表現するための代表できるフォーマットで、機械学習フレームワーク間のモデルの相互運用を目指している。 ONNXは推論に特化しているため、Pytorchなどの学習に最適化されたフレームワークで作られたモデルをONNXに変換し、推論に特化したONNX RuntimeやTensorRTなどのSDKで推論を実行することで推論を高速化することができる。
18	パイプライン構造	処理要素を直列に連結し、ある要素の出力が次の要素の入力になるようにした構造のこと。コンピュータで命令が実行されるためには、(1)命令の読み出し、(2)命令の解釈、(3)命令の実行、(4)処理結果の格納の4つのステップを経る必要がある。一つの命令を実行し、次の命令を実行するためにこの4つのステップが完了してからでは効率が悪いため、最初の命令の(1)が終わり、(2)を実行すると同時に次の命令の(1)を実行するようにした構造。
19	PIM	Processing-in-Memoryの略。メモリ回路内にプロセッサを組み込み、データの読み出し中に演算を行う回路のこと。 AI演算のように演算処理量が膨大になると、従来のメモリとプロセッサがバスで接続された構造ではメモリーとプロセッサの間のデータ移動に必要な電力が高くなるという課題があり、PIMはその解決方法の一つである。
19	HBM2	HBM(High Bandwidth Memory)の後継のメモリ規格。HBMはひとつのダイあたりの容量が2GBだったのに対し、HBM2では8GBが基本となる。積層数もHBMでは4段、HBM2では8段までと拡大されている。

【用語集】(2/2)

スライド#	用語	意味
20	Tensor命令	テンソル演算のための専用命令のこと。機械学習の主となる演算で使われる積和演算のデータは3次元以上のデータ構造をもち、テンソル(Tensor)と呼ばれる。
20	演算器アレイ	演算器をアレイ状に配置した回路のこと。 機械学習では、積和演算が繰り返し実行される。この積和演算の処理能力を向上させるために積和演算器をアレイ状に配置し、同時並行に処理を実行する。
20	マルチコア	一つのCPUパッケージ内に複数のプロセッサがある状態のこと。現在のPCに搭載されているCPUはほとんどがマルチコアになっている。マルチコアにすることで、アプリケーションを同時並行して実行することができる。
21	データフロー型	データが処理要素を移動しながら順次処理が行われるアーキテクチャのこと。データ駆動型ともいう。ノイマン型コンピュータは、CPUが処理を、メモリが命令とデータを格納し、メモリからCPUに命令が読み出され、データが読み出され、処理され、再びメモリに格納されて処理が進む。データ駆動型のコンピュータでは、データが来たら処理を行い、終わったら次のユニットに送り出して終了するユニットが、特定の処理のための構造で並んで処理が進められる。
23	PyTorch	Facebookの人工知能研究グループAI Research labによって開発されたPythonのオープンソースの機械学習フレームワークのこと。PFNは自社の機械学習フレームワークであるChainerの開発を終了し、PyTorchに移行している。
23	TensorFlow Lite	Googleが開発したオープンソースの機械学習フレームワークのこと。推論用のフレームワークで、AndroidやiOSで実行可能。学習はTensorFlowで行う。
24	Depth-Wise/Point-Wise畳み込み	畳み込みニューラルネットワークの計算コスト削減のためのアーキテクチャのこと。畳み込みニューラルネットワークの計算量は膨大で、エッジ端末では実行に時間がかかり、実用性に乏しい。そこでエッジ端末の非力な処理能力でも畳み込みニューラルネットワークを実行できるように、大幅な計算コスト削減を図るために考案されたのがDepth-Wise/Point-Wise畳み込みである。従来の畳み込み層の処理を、畳み込みフィルタの処理とチャンネル間の重み付き和の処理に分け、前者をDepth-Wiseと、後者をPoint-Wiseと呼ぶ。 計算コストが高いDepth-Wiseのフィルタを一種類にすることで幅や高さ（空間方向）の演算を大幅に減らしたことになる。

2-1-4. Arm DevSummit 2021

目次 (1/2)

1. エグゼクティブサマリ
2. Arm DevSummit 2021概要
3. プログラム
 - 3-1. 基調講演
 - 3-2. FEATURE SPOTLIGHT TALK
 - 3-3. テクニカルセッション
 - 3-4. イベント：パートナー セッション
 - 3-5. ARM エコシステム セッション
 - 3-6. APAC 固有セッション
4. 全体および注目セッション
 - 4-1. 基調講演及び全体を通してのまとめ
 - 4-2. AI
 - 4-3. SystemReadyの最新情報
 - 4-4. Arm Virtual Hardware
 - 4-5. HPCとエッジAIにおけるNVIDIAとの関係
 - 4-6. AVCC (Autonomous Vehicle Computing Consortium)

1. エグゼクティブサマリ (1/2)

1-1. 概要

- Arm DevSummitは、Armを使うDeveloper向けのイベントであり、Armの発表に加えて、Developerからの発表も多数あった。
- 昨年のDevSummit 2020以降、3月に**Armv9**アーキテクチャ、4月に**Neoverseの新コアV1,N2及びCMN-700**、6月に**CCA (Confidential Compute Architecture)**、そして、本イベントに合わせて**Arm Total Solutions for IoT**を発表した。
- Armは、チップベンダと競合しないビジネスモデルと強力なエコシステムを武器に、ターゲット市場で最も強いチップベンダをパートナーとして、その市場を押さえる戦略で発展してきた会社である。Armの成功要因は下記の通り。
 - ✓ 製造と切り離れた
 - ✓ 命令セットに理論的な美しさを求めず顧客の要求を取り入れた
 - ✓ インテルが注力しなかった携帯の市場で大きいトップになり大きな市場を取れた
 - ✓ エコシステム・コミュニティを重視、一人勝ちを目指さない戦略をとった
- Armの主な市場として従来は、携帯、組込機器だったが、今は、**AI, HPC, エッジコンピューティング, robotics, IoT, 5G**にフォーカスしている。
- インフラに組み込まれたチップが相互通信・協調動作することが当たり前になってきたため、チップの評価項目として、面積・消費電力に加え、セキュリティが大きな要素となった。Armの新しいアーキテクチャーArmv9の特徴の一つはセキュリティ。
 - ✓ サイバー犯罪は、2021年に世界で合計6兆米ドルの損害を与えると予測され、Cybersecurity Venturesは、世界のサイバー犯罪コストが2025年までに年間10.5兆米ドルに達すると予測している
 - ✓ ハッカーの攻撃にAIが使われるようになり、防御はますます困難になっている
- ArmとNVIDIAの関連の発表は4件 + Tech Hub3件。他のパートナーとの連携の発表と比べて特に多いという訳ではないが、HPCとエッジAIでの補完関係・連携はかなり強力である。

1. エグゼクティブサマリ (2/2)

1-2. ServerReadyの最新情報

- PC/サーバ市場において、かつてIntelは、プロセッサ単体のスペックでは負けながらも、強力なエコシステムによってトータルソリューションで勝ることで、RISCの挑戦を退けた。
- Armは、同等レベルのエコシステムを構築して、消費電力などのスペックで勝る戦略で、そのIntelと戦おうとしている。
- かつて最強であったIntelのコアやプロセスも、もはや最強とは言えず、コアとプロセスで最強なパートナーと組める点も、Armに有利である。
- 上記戦略の一環として、2018年に発表したServerReadyをサーバ以外に拡張したSystemReadyを昨年発表したのが、今回その後の発展を披露した。

1-3. Arm Virtual Hardware Target (VHT)

- Total Solutions for IoTについてはArmのニュースリリースや関連記事の方が詳しいが、VHTについては、その中核技術として本イベントで発表された。
- サーバ市場は、少数の強力なサーバベンダと、多数のサーバ向けSWのベンダによる市場であり、OSとそれより上位レイヤーのSWはどのHWでも動くように、エコシステムが構築されている。
- 一方、IoT市場は、ロングテールといわれる非常に多品種のHWが必要な市場であり、HWとその上で動くSWを開発して製品システムに仕上げる。
- VHTはHW開発と並行して効率的にSWを開発するための技術であり、これによって開発期間を5年から3年に短縮できた例が紹介された。

1-4. RISC-Vとの関係

- Armの強力なエコシステム構築にはコストがかかり、独占状態もあって、割高感が出て、スモールビジネスには敷居が高い。
- これに対して、RISC-Vのエコシステムはオープンコミュニティの力で構築されており、Western Digitalの積極的採用もあり、フリーでありながら、急速に充実してきている。
- スモールビジネスにとっては既にArmより有効であるが、Armもエコシステム強化や強力なパートナーの取り込みを進めており、RISC-VがArmをどこまで脅かすのか、今後注目したい。

2. Arm DevSummit 2021概要 (1/2)

1. 開催日時

2021年10月19日 - 21日 (3日間)

2. 開催形式

完全バーチャル・イベントで開催

2021年12月8日までコンテンツの閲覧可能

3. 構成

全140以上のテクニカルセッション、オンラインワークショップ、パートナーセッション、Tech Hub、参加者同士の交流の場などからなる。

※今年のArm DevSummitのコンテンツは7つのカンファレンストラックに分類されている。次スライドに7つのトラックを示す。また、プログラムにはトラックとの対応付された発表にはトラック番号を記載した。



URL <https://devsummit.arm.com/en>

2. Arm DevSummit 2021概要 (2/2)

No	トラック	内容
1	エンドポイント、エッジ、クラウド向けのクラウドネイティブ	ハイパースケールデータセンター、5Gネットワーク、エッジにArmテクノロジーが浸透する中、ソフトウェア開発者は、Armベースの環境に対応したOS、ツール、コンテナや仮想化のための最新言語サポートの準備しています。ここでは、エコシステム全体でArmアーキテクチャ対応のクラウドネイティブ開発を促進するベストプラクティスをご紹介します。
2	モビリティの未来はソフトウェアで決まる	自律走行技術と電動化の進展は、自動車業界におけるモビリティの未来を変革し、より安全、効率的で、手頃なソリューションが世界中で実用化されています。そして車載システムは単純なディスクリートECUコンポーネントを利用したソリューションから、異種のテクノロジーをSoCやサブシステムに組み合わせたものになりつつあります。これに伴い、ソフトウェアの複雑性は増加してきています。同様の傾向は産業用ロボットなどの類似分野にも見られます。クラウドネイティブの設計は、この複雑性を抑え、リアルタイム性や機能安全の要件を満たすソフトウェア定義の車載ソリューションの大規模導入に役立ちます。このトラックでは、ソフトウェア定義車両への道を歩む開発者に有益な情報や知識を紹介します。
3	インタラクティブなコンシューマエクスペリエンスの開発	機械学習の急速な進歩、拡張現実の用途の拡大、ソーシャルコンテンツやモバイルゲームの急成長など、モバイル機器やノート型PCの開発者にとってこれほど活気ある時代はありません。個人データのセキュリティ確保も再び注目を浴び、まだまだ多くの学習と検討が必要です。ここでは、業界トレンドに関する情報や実践的なワークショップに加え、Armをベースとした開発のパフォーマンスを高める-最先端の開発手法やツールをご紹介します。
4	実用的なIoT向けのソフトウェア開発	IoTは、接続性、セキュリティ、効率的なデータ処理と取り扱いなど、多くの課題を残しつつ、普及と大規模導入という新しい段階に入りました。しかし、機械学習対応エンドポイントの登場、主要機能の標準化を推進する動き、基本的なテクノロジーの成熟により、開発をスピードアップする道が開けました。このトラックでは、最新ツール、トレンド、テクノロジーに関する知識をソフトウェア開発者に提供し、プロトタイプから生産への時間短縮をお手伝いします。
5	AIの実用化：開発から運用まで	データ量の増大とソフトウェアの革新により、テクノロジー界ではAIと機械学習が注目の的ですが、複雑なニューラルネットワークを駆使した世界的な感染症対策から機械学習でルービックキューブを解くロボットまで、次々と新しいことが可能になりつつあります。機械学習のエンジニアやテクノロジー専門家と交流し、AIの未来に向けた開発作業を実際に体験してみてください。
6	ハードウェア設計手法	現代のチップ設計はシリコンで終わりません。セキュアで消費電力の低い複雑なマルチコアシステムに対する需要が高まるとともに、設計者は新しいプロセッサやオフロード対策でターゲットを絞った効率的なソリューションの開発に努めています。ここでは、モデリングやツールの進歩がどのようにハードウェアとソフトウェアの協調設計を促進するか、そしてEDAや設計サービスが実績あるチップ/プラットフォームベンダーの新規市場参入にどのように役立つかを解説します。
7	HPC、クラウド、5Gインフラに向けたArmテクノロジーの最適化	ハイパフォーマンスコンピューティング（HPC）やクラウドや5Gインフラストラクチャでは、基盤となるハードウェアやソフトウェアの性能を引き出すために、広範囲にわたる最適化が必要です。ぜひ、ワークロードのバランス、マルチスレディング、あるいはよりコスト効率の高いパフォーマンスや分散データストレージと高帯域幅ネットワークのテクニクについての知見をご確認ください。また、ArmがArmエコシステムにとって業界標準をどのようにサポートするかなど、ツール、ソフトウェア、および開発プラットフォームに関するエコシステムの概要もご紹介します。

3. プログラム

3-1. 基調講演

講演タイトル	講演者/所属	対応トラック
パフォーマンスとパーパス：これからの世界に必要な不可欠となる新たなコンピューティング	Simon Segars Chief Executive Officer / Arm	-
システムを考慮したデザイン	Mohamed Awad Vice President, IoT & Embedded / Arm	-
次世代Windows on Armの開発	Hari Pulapaka Partner Group Program Manager / Microsoft	3
神経の可塑性を超えて：AI、生物学、テクノロジー	Ian Bratt Fellow and Senior Director of Technology, Machine Learning Group / Arm	-
テクノロジーを語る	Rene Haas/President, IP Products Group (IPG), Arm Suraj Gajendra/Senior Director, Technology Strategy, A&I, Arm Andrew Rose/Chief System Architect and Fellow, Arm Mark Hambleton/Vice President, Open Source Software, Arm	-
AWS & Arm - クラウドからエッジを目指す開発者を支援	Dipti Vachani/SVP GM Automotive & IoT, Arm Bill Foy/Director, World-Wide Automotive Go-to-Market, AWS Vin Sharma/GM and Director, ML Engines and Edge ML, AWS AI, AWS Raj Pai/Vice President, Product Management, Amazon EC2, AWS	1
オープンソースソフトウェアと次の10年のイノベーション	Andrew Wafaa / Distinguished Engineer, Senior Director Software Communities, Arm Steve Geary / Director, Open Source, Wind River Nuritzi Sanchez / Senior Open Source Program Manager, GitLab, Inc. Stormy Peters/Director, Open Source Programs Office, Microsoft	-

3-2. FEATURE SPOTLIGHT TALK

講演タイトル	講演者/所属	対応トラック
バーチャルリアリティにおけるデータエンジニアリングのための環境整備に関する考察	Dr. Newton Campbell Senior Principal Solutions Architect / SAIC/NASA	-
自然界のモノのインターネット：音響、AI、保護	Topher White Chief Executive Officer and Founder / Rainforest Connection (RFCx) Chrissy Durkin Director of International Expansion / Rainforest Connection	-
地球にやさしいハイパフォーマンス・コンピューティング	Brett Bibby Senior Vice President, Product / Unity	3

3-3. テクニカルセッション（1/3）

講演タイトル	講演者/所属	対応トラック
IoTのエンドノート開発に対応するCI/CD、MLOpsワークフロー	Reinhard Keil Senior Director of Embedded Tools / Arm	4
SystemReadyの最新情報	Dong Wei Standards Architect and Fellow / Arm	6
クラウドネイティブのスーパーコンピューティング：ベアメタルのSecured Supercomputing Architecture	Gilad Shainer Senior Vice President Networking / NVIDIA	7
エッジでのコンフィデンシャル・コンピューティングに対応するエンクレーブ・デバイスの設計	Eustace Asanghanwa / Principal Program Manager, Microsoft Paul Howard / Principal System Solutions Architect, Arm Femi Idowu / Senior Cloud Solutions Architect , Microsoft Evert Pap / Principal System Solutions Architect, Scalys	1
Arm版Windows開発の基礎	Marc Sweetgall Senior Program Manager / Microsoft	3
Oracle Cloud Infrastructure (OCI) の第一歩	Peter Pouliot Developer Advocacy / Ampere Computing Manish Kapur Director, Product Management / Oracle Sanjay Pillai Product Manager / Oracle	1
次世代モバイルネットワークに対応する大規模ネットワーク	Phillip Ritter Magma Community Evangelist / TeamRitter LLC, consulting for Facebook Shaddi Hasan Assistant Professor of Computer Science / Virginia Tech	7

3-3. テクニカルセッション（2/3）

講演タイトル	講演者/所属	対応トラック
シートベルトを締めて：ソフトウェア定義車両の到来	Robert Day / Director, Automotive Partnerships, Arm Stefano Marzani / Principal Specialist SA, Autonomous Driving, Amazon Web Services Pierre Olivier / Chief Technology Officer, LeddarTech Martin Schleicher / Head of Software Strategy, Continental	2
TinyML予測メンテナンスソリューションのローコード開発	Massimo Banzi Co-Founder / Arduino SA	5
MTEを使用してAndroid 12でメモリの安全性を向上	Serban Constantinescu Product Manager / Google	3
System Ready IR認証の価値と経路	Olivier Bernard Director, High Performance IoT / Arm Grant Likely Senior Technical Director, Software Architecture / Arm	4, 6
IoTロングテールに対応したクラウドベースのソフトウェア開発を実現	Stefano Cadario Director, Software Product Management / Arm	4
Autowareがクラウドネイティブの自動運転システム開発を定義	Shinpei Kato Chief Technical Officer / Tier IV Girish Shirasat Chief Software Architect / Arm	2
Maliで簡単にグラフィックスパフォーマンスを向上	Peter Harris Technical Product Director / Arm	3

3-3. テクニカルセッション (3/3)

講演タイトル	講演者/所属	対応トラック
エッジ上のエンドツーエンドのクラウドネイティブ・コンピュータビジョン・アプリケーション	Megh Makwana Solutions Architect / NVIDIA Ajeet Raina Developer Relations Manager / Redis Labs	1
リアルタイムゲームエンジンにおけるニューラルネットワークの実行	Alexandre Ribard Machine Learning Engineer / Unity	3
HPCのバランスに革新 (再び) - クラウドの面から	Brendan Bouffler Head of HPC Developer Relations / Amazon Web Services	7
車載サイバーセキュリティ：信頼できる車両アーキテクチャ・エコシステムに向けて	John Moore Vehicle Security Architect/ Tech Specialist / Ford	2

3-4. イベント・パートナー セッション (1/2)

講演タイトル	講演者/所属	対応トラック
Arm CoreLink CMN-700ベースのSoCおよびシステムのシステムレベル検証の高速化	Nick Heaton Distinguished Engineer / Cadence David Koenen Senior Product Manager of Interconnect IP / Arm	7
Armベースのアプリケーションを時代に合わせて変化させる	Shawn A. Prestridge US FAE Team Manager / IAR Systems, Inc.	4
Oracle Ampere A 1、クラウド開発者に新たな機会を提供	Ross Brown / VP, OCI Marketing, Oracle Jeff Wittich / Chief Product Officer, Ampere Matt Leonard / VP, Oracle Compute, Oracle Dan Gerrity / SVP, Oracle Developer Ecosystem, Oracle	1
インタラクティブ・インテリジェンスの進化	Mohammed Dogar Senior Director Global Business Development / Renesas Electronics Manny Singh Principal Marketing Manager / Renesas Electronics	3, 4
Siemens EmbeddedとMendix - 機能豊富なIoTソリューションを迅速に実現するためのレシピ	Muhammad Shafique Senior Product Manager / Siemens	4
Arm上のCUDA向けHPC開発を簡素化する最新Nsight開発ツール	Jackson Marusarz Product Manager, Developer Tools / NVIDIA	7
オープン・プラットフォーム・スタンダードでオートモーティブ・エコシステムの成長を実現	Jered Floyd Technology Strategist, Office of the CTO / Red Hat	2

3-4. イベント・パートナー セッション (2/2)

講演タイトル	講演者/所属	対応トラック
NanoEdge AI Studio - エッジAIソリューションを迅速かつ効率的に開発するためのパワフルなソリューション	Louis Gobin Field Application Engineer / STMicroelectronics Markus Mayr Product Marketing Manager / STMicroelectronics	4
あらゆる形態のコンピュートを実現：アームとシノプシス社、そして民主化されたHPCの登場	Brian Millar Principal Engineer / Synopsys	7
プレシリコンで加速するArmv9ベース・システムの安全性とセキュリティの検証	Maxwell Hinson Lead Technical Marketing Engineer / Green Hills Software Frank Schirrmeyer Senior Group Director, Solutions & Ecosystem / Cadence Design Systems	6
Arm向けSUSEオープンソース・ツール・エコシステムの紹介	José Betancourt Director, Solution Partners & Alliances / SUSE	1

3-5. ARM エコシステム セッション

講演タイトル	講演者/所属	対応トラック
神話の崩壊へ：ORAN、COTS、CloudNativeをベースにした超効率的な大規模MiMo 5 Gプラットフォーム	Kuntal Chowdhury / SVP & GM, AI & Analytics, Mavenir Soma Veyalutham / General Manager, AI and 5G, NVIDIA Panch Chandrasekran / Director, Carrier Infrastructure Segment Marketing, Arm Henry Justin Calvert / Senior Director, Head of Future Networks, GSMA	7
炭素排出量の削減へ：SoC開発から製品展開、そしてエンド・オブ・ライフまで	Hannah Peeler Graduate Researcher / Arm Joshua Randall Staff Research Engineer / Arm Hilary Tam Sustainability Strategy Director / Arm	7
Ampere® Altra®上で最大のスケールビリティに最適化を実現するCanonical社のAnbox Android in the Cloud	John O'Neil Principal Software Engineer / Ampere Computing Simon Fels Engineering Manager / Canonical	1
テイルズ・フロム・エッジ：Project Cassiniとデータセンターを超えたクラウドネイティブの世界	Mark Abrams Field Engineer Edge Specialist, SUSE Amar Padnamabhan Member, Board Of Directors, FreedomFI Henri Parmentier Senior PM, Adlink Embedded platforms & modules, Adlinktech Augustine Nebu Philips Director, Segment Marketing for Edge Compute and Enterprise Networking , Infrastructure Line of Business, Arm	1

3-5. APAC 固有セッション

講演タイトル	講演者/所属	対応トラック
新時代のBYD：チップ半導体技術をベースにどのように電子化、インテリジェント化するか	Han Bing President Assistant and Director of Auto Product Planning & New Technology Research Institute / BYD	2
VISION-S プロジェクト：ソニーのモビリティに対する取り組み	川西 泉 常務 AIロボティクスビジネス担当 / ソニーグループ株式会社	-
DEEPXが実現するAIoT時代に向け	Lokwon Kim CEO / DEEPX	5
テンセントにおけるArmサーバー・アーキテクチャの実践	Li Chengdong Hardware & Software Co-optimization Architect / Tencent CSIG	1, 7
KDDIによる5G+XRを活用したDXの取り組みについて	上月 勝博 パーソナル事業本部 サービス統括本部 5G・XRサービス企画開発部 / 部長 / KDDI株式会社	-
ポストCOVID-19時代のエッジコンピューティングの革新と応用	CK Tseng / President, Arm Taiwan Jordan Wu / President, CEO and Director, Himax Technologies Victoria Lee / Senior Partner Manager, Arm Taiwan Jerry Kao / COO & President, IT Products Business, Acer	4
Armテクノロジーと歩むテレチップスの現状と将来	Leanne Lee VP/Head of Future Strategy Group / Telechips, Inc.	-
エッジAIでどう変わる？世界の今と今後	David Hsu / FAE Director, Arm Taiwan Odin Shen / Principal Field Application Engineer, Arm Taiwan Hsien-Hsin Sean Lee / Area Research Lead, Facebook AI, Facebook, Inc. Wei-Fen Lin / Professor, School of Computing, Miin Wu School of Computing, National Cheng-Kung University	1

4. 全体および注目セッション

4-1. 基調講演及び全体を通してのまとめ（1/4）

- 基調講演:パフォーマンスとパーパス：これからの世界に必要な不可欠となる新たなコンピューティング（Arm CEO / Simon Segars）
- 基調講演:システムを考慮したデザイン（Arm VP of IoT & Embedded / Mohamed Awad）

(1) 応用される分野、製品

- 携帯・組込機器・AI・HPC・エッジコンピューティング・robotics・IoT・5G

(2) 当該分野における技術ニーズ

- 高性能コンピューティング、低消費電力、AI/ML、開発期間短縮

(3) 発表（セッション）の概要

- Armの歴史を振り返り今後の戦略・注力領域を説明

(4) 技術的に特筆すべき点

- Armv9アーキテクチャ、
- Neoverseの新コアV1,N2及びCMN-700CCA (Confidential Compute Architecture)、
- Arm Total Solutions for IoT

(5) 本発表（セッション）が注目される理由

- Armの全体的戦略をカバーする講演である

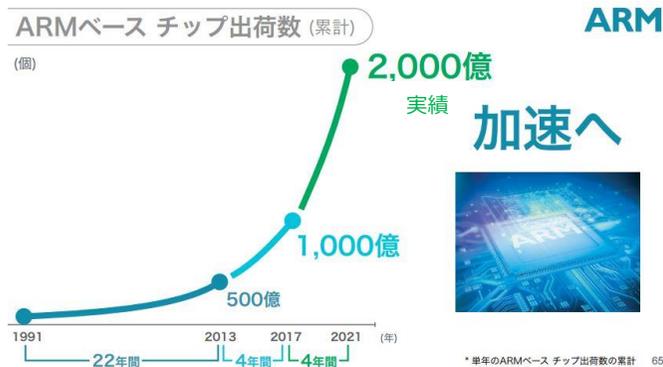
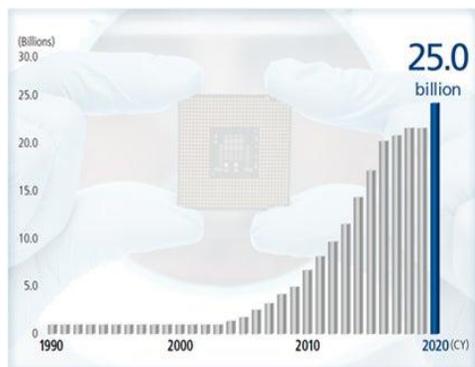
(6) 実用化までの課題

- 特になし

4-1. 基調講演及び全体を通してのまとめ (2/4)

■ Armに関して

- Segars, Awadの基調講演では、**累計2,000億個以上**、2020年には**250億個以上**のチップを出荷、と言及



今回の基調講演では、言葉での言及はあったが適当なスライドがなかったので、ANNUAL REPORT 2021 Arm: Business Strategy、2017年3月期のSoftBank2017決算説明会資料から引用
予想通り出荷数が伸びていることがわかる

- アーキテクチャーのラインアップは、
 - Cortex-A, Cortex-R, Cortex-M (世代で分けるのではなく、用途別にA:スマートフォン向け、R:リアルタイム制御向け、M:低コストなマイコン制御向け、と分けている)
 - Mali (GPU), Ethos (NPU)
 - Arm v9(2021年3月発表、最新アーキテクチャー)
- パートナー数: 1,000社以上、開発者数: 数百万人のエコシステムを形成

4-1. 基調講演及び全体を通してのまとめ (3/4)

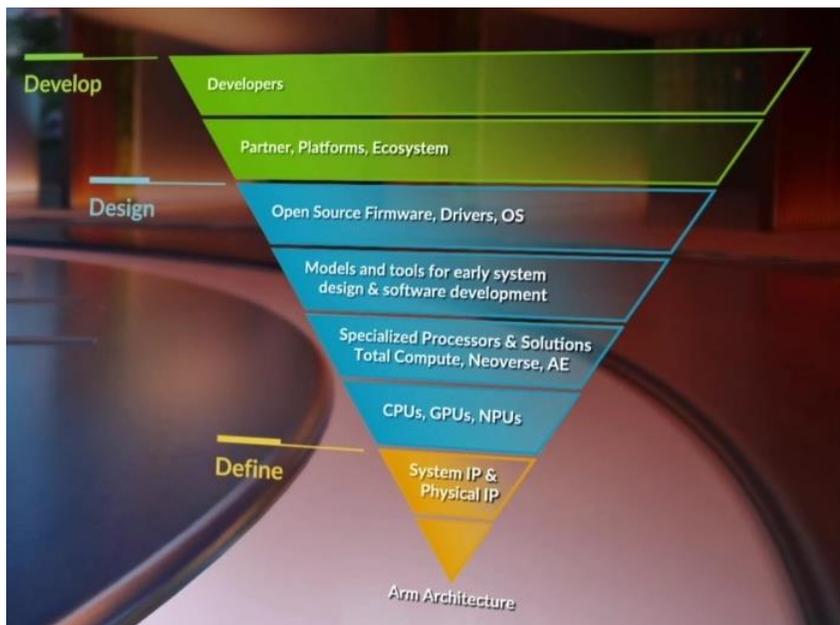
Arm ロードマップ

	シリーズ	応用分野	Production	Now (2021)	2022+	
CPU	Cortex-M	低コストなマイコン制御向け	Armv8 M55 (TSMC22ULL)	Armv9		
	Cortex-R	リアルタイム制御向け	R82			
	Cortex-A	スマートフォン向け	A58 A78	A510 (A55+35% uplift) A710 (A78+10% uplift)		
	Cortex-X	カスタム	X1	X2 (X1+16% uplift)		
			7nm	7/5nm	5nm	
	Neoverse E	ネットワーク・データプレーン用プロセッサ、低消費電力5Gゲートウェイ	E1 A53+2.7x throughput	Armv8	Poseidon	
	Neoverse N	クラウド・エッジ・5G	N1 A72+60% ST perf. uplift		N2 N1+40% St Perf. Uplift SVE 2x128b,bFloat16	+30% infra WL perf. ML/Vector uplift Greater Core density
	Neoverse V	データセンター、HPC、AI/ML	Armv8+SVE	V1 N1*50% St Perf. Uplift SVE 2x256b,bFloat16	Armv9	
GPU	Mali	グラフィック、ゲーム	Mali-G76 Bitfrost3	Mali-G710 Valhall3	2022 GPU G76x4.7 FP32 ML perf.	
NPU	Ethos	エッジAI	U55 (TSMC22ULL)	U65		

4-1. 基調講演及び全体を通してのまとめ（4/4）

■ Armの基本戦略

- 市場が伸びていくと予想される代表的なアプリケーションエリアを絞り、エコシステムのパートナーへのソリューション提供に投資。Awadの基調講演の図が階層を表している（下図）。
- 従来は携帯、組込機器だったが、今は、**AI, HPC, エッジコンピューティング, robotics, IoT, 5G**にフォーカス
⇒しかしArmが全部をやるということではなく、Armが得意なところで価値を提供 * 事例を4-2以降
 - ・低電力・省エネルギー（二酸化炭素軽減）
 - ・ソフトウェア・ハードウェア協調デザイン（開発期間短縮）
 - ・複数チップ間の協調動作のための通信時のセキュリティ（Confidential Computing）



CCA(Confidential Compute Architecture)



通常のプログラムはRealmという暗号化された領域に格納され、OSおよびハイパーバイザーから分離される

<https://www.arm.com/ja/why-arm/architecture/security-features/arm-confidential-compute-architecture>

4-2. AI (Ian Brattの基調講演及びテクニカルセッション) (1/2)

- 基調講演:神経の可塑性を超えて : AI、生物学、テクノロジー (Arm Fellow and Senior Director of Technology, Machine Learning Group/ Ian Bratt) およびAI/ML関連テクニカルセッション

(1) 応用される分野、製品

- 分野 : AI、
- 製品 : 携帯・組込機器・HPCサーバー・エッジコンピューティング・robotics

(2) 当該分野における技術ニーズ

- AI/ML、高性能コンピューティング、低消費電力、開発期間短縮

(3) 発表 (セッション) の概要

- AIの進化について、人の発達段階がセンサー(感覚)→言語→認知と能力を獲得することと類似して、AIも今後、認知の領域での研究が進む。
- ArmのAI/MLに対する取り組みを製品ラインアップ、パートナーとの事例を通して紹介

(4) 技術的に特筆すべき点

- Armv9 on CPU、Mali GPU、Ethos NPUにおけるAI/ML対応命令、Arm NNライブラリ
- ニーズに合わせた広いラインナップ&ソリューションの提供、低消費電力をArmの強味としている。

(5) 本発表 (セッション) が注目される理由

- ArmのAI/MLに対する取り組み、戦略をカバーする講演である。

(6) 実用化までの課題

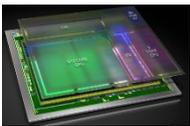
- 認知の領域に適した計算手段と命令セットにどのようなものがでてくるか注目したい。

4-2. AI (Ian Brattの基調講演及びテクニカルセッション) (2/2)

- ArmがAIすべてをやる&強いわけではない→ 強味をしっかり発揮
- ニーズに合わせた広いラインナップ、ソリューションの提供が強み

大手の独自AIチップには使いやすい、CPUコアを提供

エッジ、HPCなどには多様なIPコア、ライブラリを提供

	Pc向けSoC	携帯向けSoC	自動運転向けSoC
	Apple M1 Max	Google Tensor	Nvidia Xavier
			
CPU	Armベース 高性能コアFirestorm x8 高効率コア Icestorm x2	高性能Cortex-X1 x2 + 高効率コアの 「2+2+4コア」	Carmel ARMv8.2 *
GPU	独自GPU	TPU x 20コア	Volta GPU 512コア

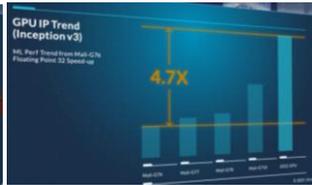
CPU
Arm v9

GPU
Mali-G76

NPU
Ethos-
U55/U65



Armv8.4/8.6で
Dot Product
Matrix Multiply
BF16
を導入
Armv9で
SVE2 を導入



次世代のMali-
G76で
x4.7倍の高速化
を実現



組込機器、IoT向け
Alif のマルチコアデバイ
スは、Cortex-M55と
Ethos-U55を統合、
Cortex-M55のx75
の性能
Relajet 音声NN
付イヤフォン

Cortex-M



Arm NN Library

DevSummitで力を入れて宣伝

DevSummitでは言及なし

*次世代のOrinでは
Cortex-A76ベース
のHercules CPU コ
アを採用

※CPUに高性能コアと高効率コアの両方を積む
のは最近のトレンドで、バッテリー駆動時間を延
ばすため

4-3. SystemReadyの最新情報（1/6）

講演：「SYSTEMREADYの最新情報」「System Ready IR認証の価値と経路」

(1) 応用される分野、製品

- IoTゲートウェイ、エッジサーバ、インフラエッジと5G, インテントベースネットワーク

(2) 当該分野における技術ニーズ

- OSおよびアプリケーションの汎用性

(3) 発表（セッション）の概要

- SystemReadyはOSがArmベースのサーバ、インフラエッジ、組み込みIoTシステムで「問題なく動く」ことを認証するための枠組み
- 今回は、プログラムの要件、各種テスト、認証プロセス、権利放棄、プリシリコンのコンプライアンステスト、PSA認証プログラムとの関係における最新情報を紹介

(4) 技術的に特筆すべき点

- 枠組みの適用範囲を拡大（従来技術はサーバに特化）

(5) 本発表（セッション）が注目される理由

- 前回のDevSummitで発表されたSystemReadyの最新情報が得られた。

(6) 実用化までの課題

- 既に実用レベルの技術

4-3. SystemReadyの最新情報 (2/6)

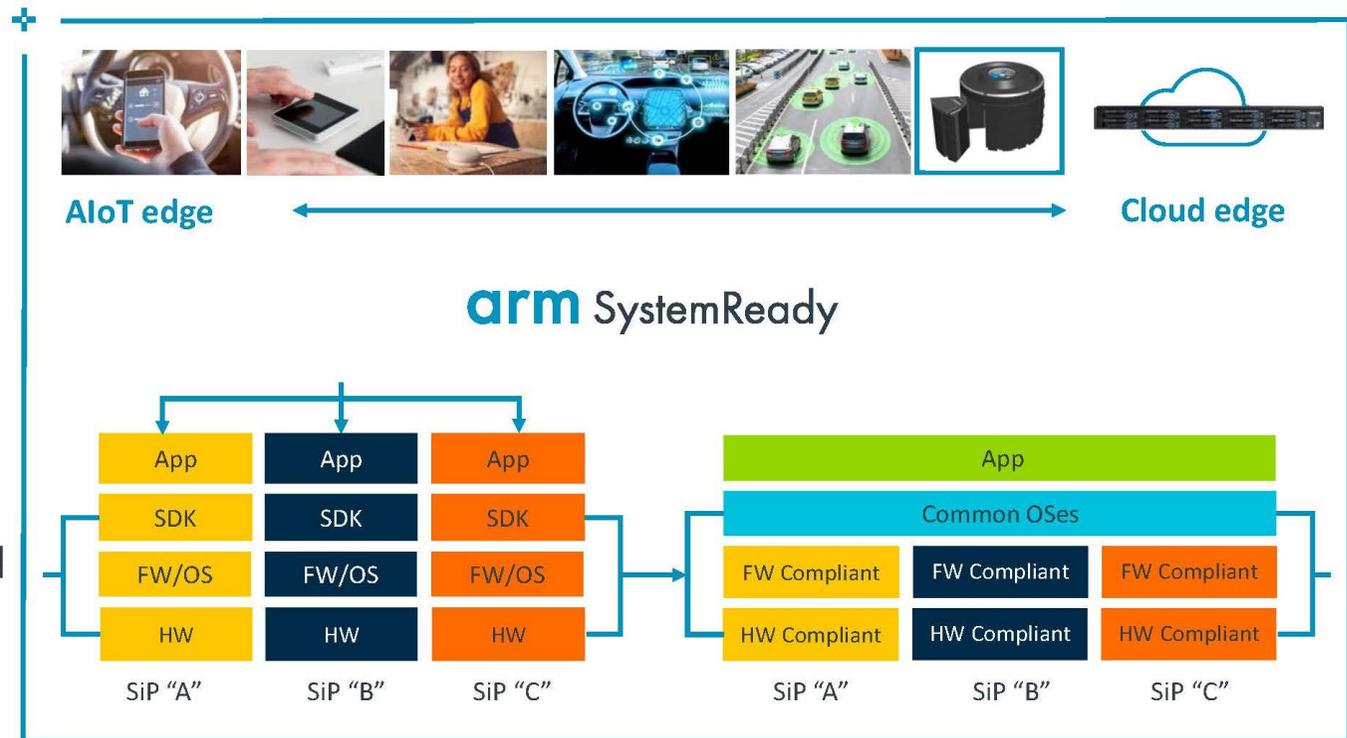
背景・Vision

- SystemReadyは、元々、x86が実現しているHW非依存のOS・アプリ開発をARMでも可能なように整備したものを、サーバだけでなく、他のシステムにも適用すべく発展させたもの

概要

- 前回のDevSummitでSystemReadyコンプライアンス認証プログラムを発表
- OSがArmベースのサーバー、インフラエッジ、組み込みIoTシステムで「問題なく動く」ことを認証するため
- 今回は、プログラムの要件、各種テスト、認証プロセス、権利放棄、プリシリコンのコンプライアンステスト、PSA認証プログラムとの関係における最新情報を紹介

Vision



Software Can Just Work on Arm-based Devices

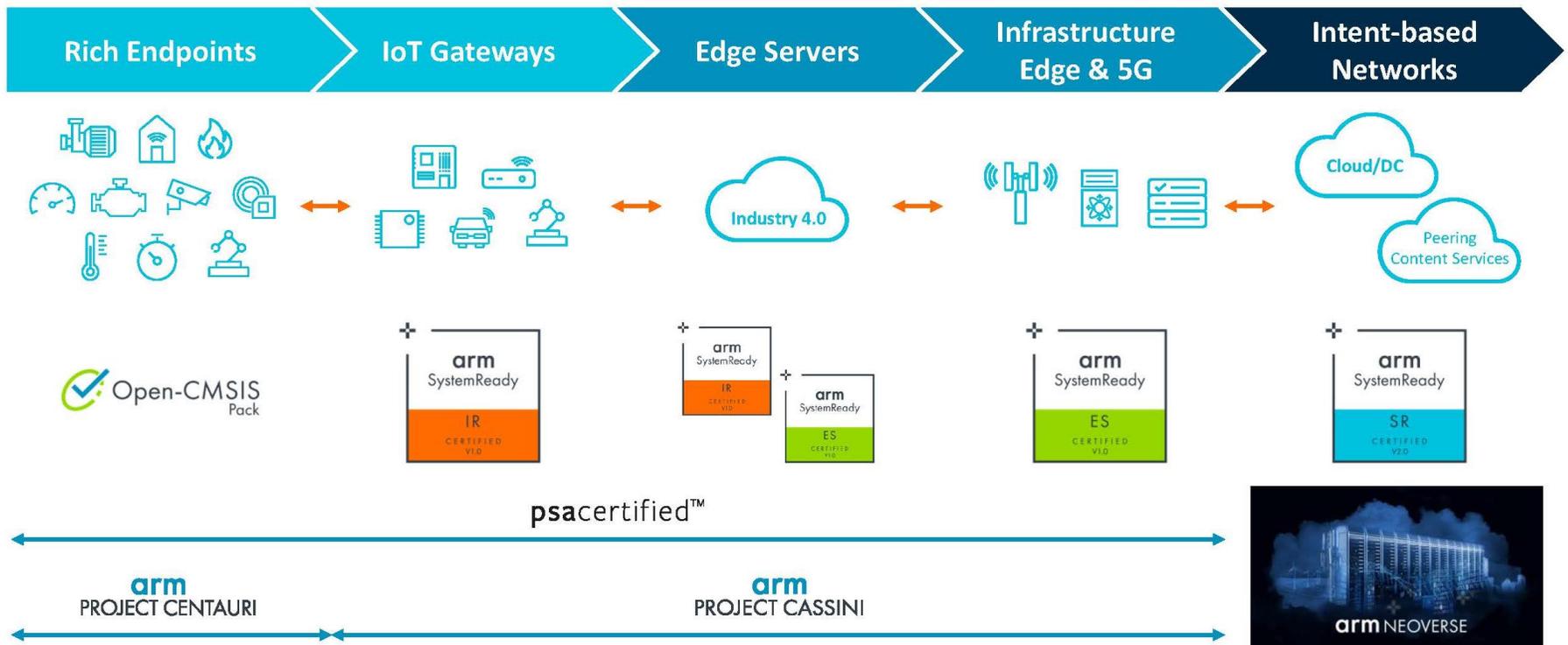
4-3. SystemReadyの最新情報 (3/6)

- SystemReady IR (IOT Ready), ES (Embedded Server Ready), SR (Server Ready) で広範な領域をカバー
- Cortex-A, Cortex-X, Neoverseが対象
- Rich Endpointsは、Cortex-Mの適用を想定し、Open-CMSIS Pack, PROJECT CENTAURIでサポート

IoT Deployment Challenges: Fragmentation



A frictionless edge is critical to unlocking value from IoT

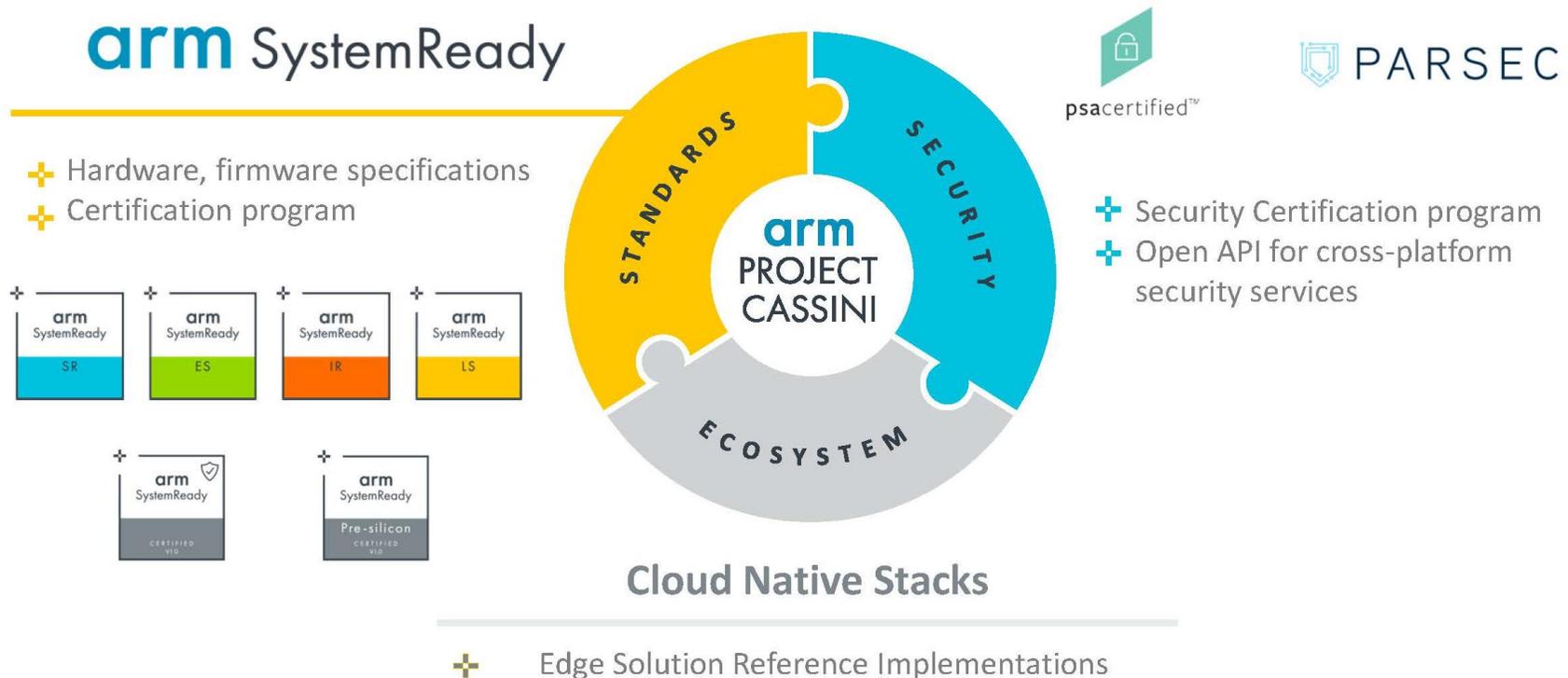


4-3. SystemReadyの最新情報 (4/6)

- **Project Cassini**によって、EdgeでのCloud Nativeシステムの実現を支援
- SystemReadyで、Hardware, Firmwareの**STANDARDS**準拠を認証
 - SR, ES, IR, LS (LinuxBoot Server Ready), certified V1.0, Pre-silicon certified V1.0で構成
- **psacertified**と**PARSEC**で、**SECURITY**認証
- Cloud Native Stacks, Edge Solution Reference Implementationsなどの**ECOSYSTEM**が充実

Project Cassini

An industry initiative to ensure cloud-native experience at the Edge



4-3. SystemReadyの最新情報 (6/6)

- 2021年9月時点で認証されているデバイス

Arm SystemReady – Certified Devices (Sept 2021)

arm SystemReady

Band	Certified Systems
	 • Mt Jade Platform  • Ampere Mt. Snow Platform  • Ampere Altra Server 2U Mt. Jade • Ampere Altra Server 2U Mt Snow NVMe
	 • LX2160A RDB • LS1046A FRWY • LS1046A RDB  MARVELL • OCTEON TX2 CN9130 DB  • HK-6010  • Raspberry Pi 4 Model B • Raspberry Pi 400  SolidRun Embedded Edge Computing • HoneyComb LX2 Workstation • MACCHIATObin Double Shot • CEx7 CN9132 Eval Board
	 • NXP i.MX8M Mini EVK  Compulab • Compulab IOT-GATE-iMX8  • Leez P710 Gateway  • Raspberry Pi 4 Model B • Raspberry Pi 400  • RockPro64  瑞芯微电子 • Toybrick TB-RK3399Pro

4-4. Arm Virtual Hardware (1/5)

講演：「IoTのエンドノート開発に対応するCI/CD、MLOPSワークフロー」

「IoTロングテールに対応したクラウドベースのソフトウェア開発を実現」

(1) 応用される分野、製品

- Rich Endpoints (Cortex-Mを使う分野・製品)

(2) 当該分野における技術ニーズ

- IoTとMachine Learning (ML) で大幅に複雑化したSW開発への対応
- Continuous Integration / Delivery or Deployment (CI/CD) への対応

(3) 発表（セッション）の概要

- Virtual Hardware Target (VHT) でSWとHWの並列開発を可能にし開発期間を短縮
- VHTの構成とその具体例のCorstone-300を説明

(4) 技術的に特筆すべき点

- HWスペックからクラウドサーバ上にVHTを構築し、実HW開発と並行に、実HWより高速なSW検証およびSW統合が可能

(5) 本発表（セッション）が注目される理由

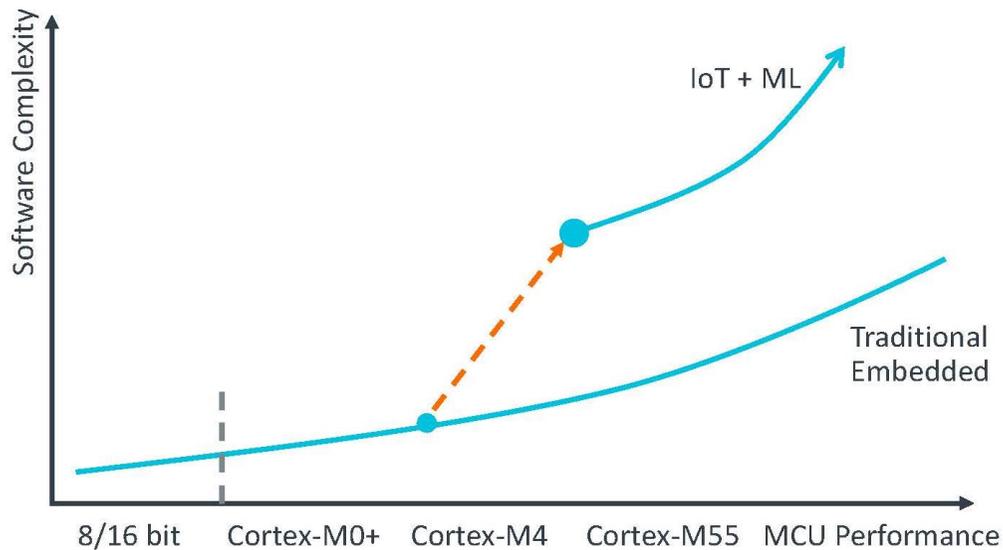
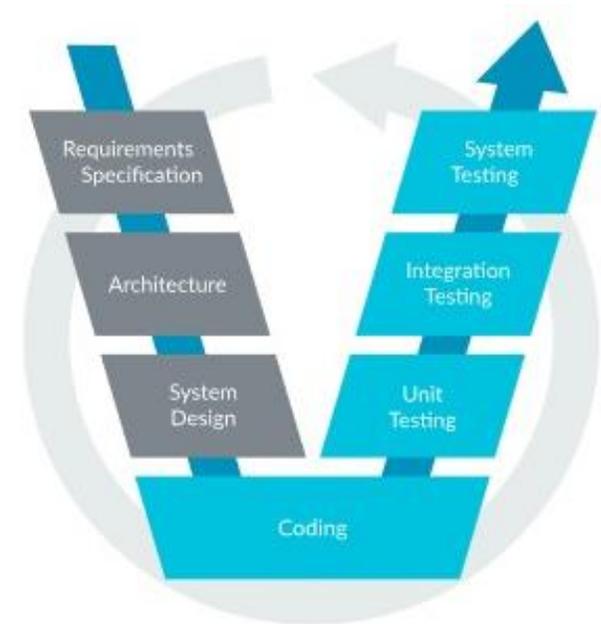
- イベントに合わせて発表されたArm Total Solutions for IoTの中核技術

(6) 実用化までの課題

- 既に実用レベルの技術

4-4. Arm Virtual Hardware (2/5)

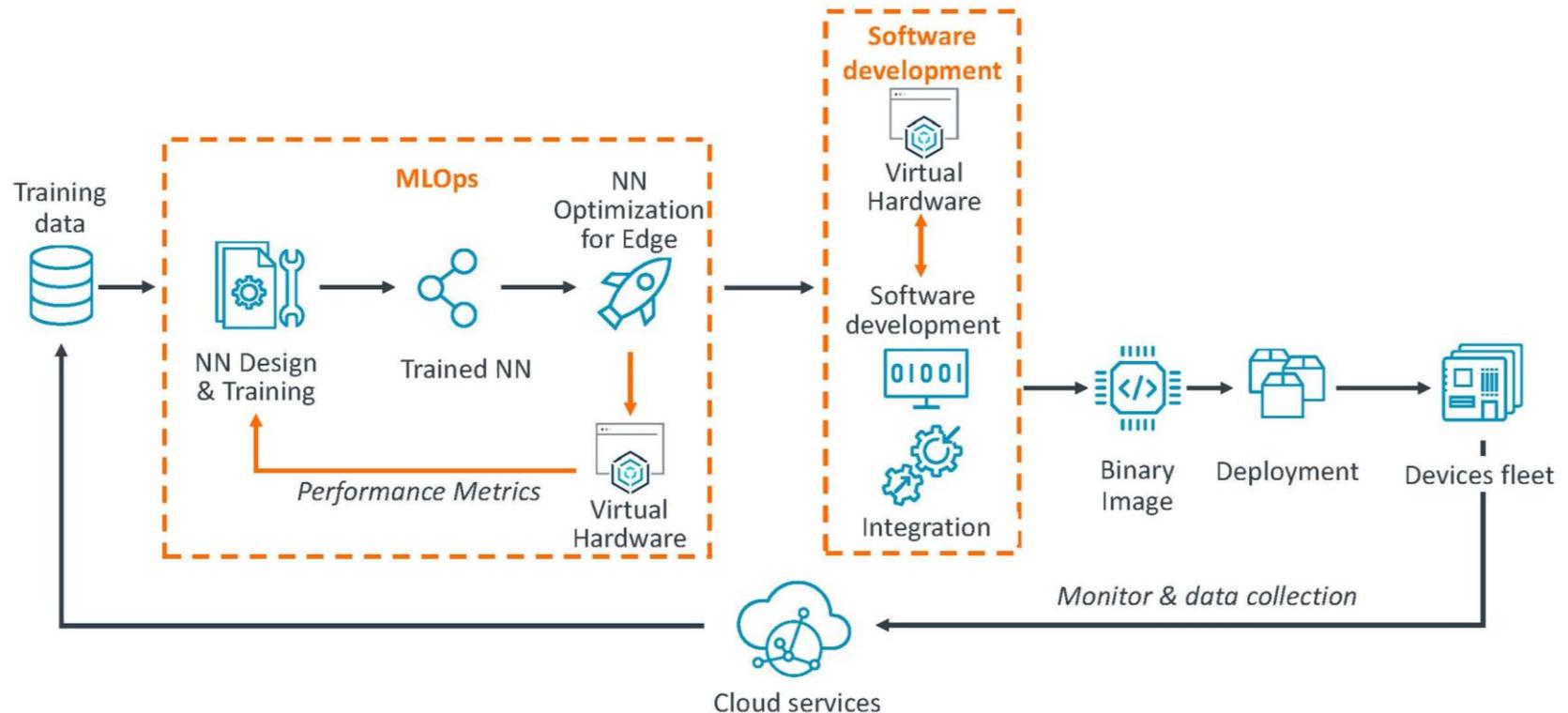
- IoTとMachine Learning (ML) で、SWが大幅に複雑化
- Continuous Integration / Delivery or Deployment (CI/CD) が開発手法の主流へ（下記を容易にする）
 - 開発時の大規模システムの統合
 - 出荷後のシステムを最新の状態に保持
- Arm Virtual Hardware Target (VHT) はイベントに合わせて発表されたArm Total Solutions for IoTの中核技術
- VHTで、SWとHWの並列開発を可能にし、上記状況にも対応



4-4. Arm Virtual Hardware (3/5)

- Arm Virtual Hardware Target (VHT) を活用した開発フロー
 - 典型的なケースで製品開発期間を5年から3年に短縮
 - HW開発と並行して、VHTでSW開発
 - MLOpsのチューニング
 - SW DevelopmentとIntegrationへの活用

An Intelligent Edge Software Development Flow



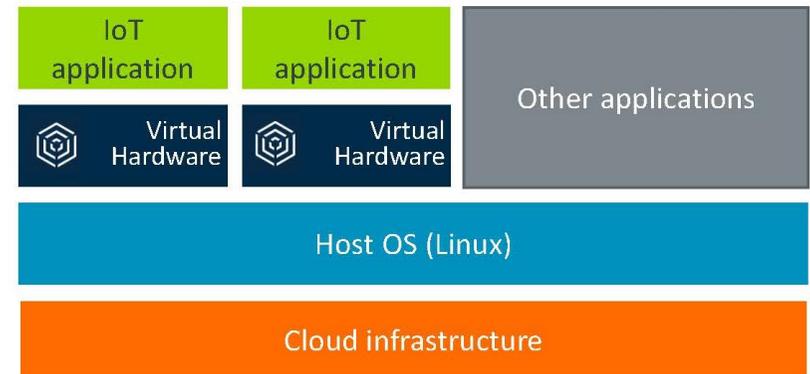
4-4. Arm Virtual Hardware (4/5)

■ Virtual Hardware の構成例

What's Arm Virtual Hardware?

Virtual Hardware Targets are the IoT equivalent of Virtual Machines

- An Arm Virtual Hardware Target is a functionally accurate representation of a physical SoC, simulating its software-visible behavior
- Runs as a simple application in a Linux environment for easy scalability in the cloud
- Remove dependency from RTL or silicon availability

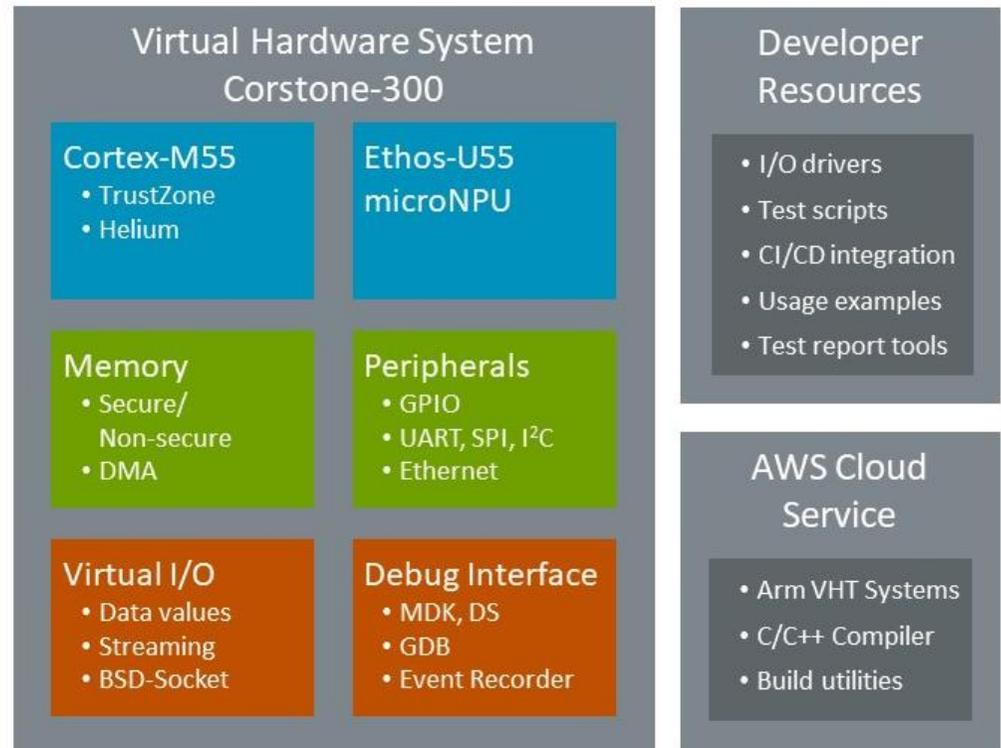


4-4. Arm Virtual Hardware (5/5)

■ Corstone-300の例

Arm Virtual Hardware

- Precise **simulation models** of Cortex-M device sub-systems designed for complex software verification and testing
- Runs any RTOS or bare metal code
- Provides virtual peripheral interfaces for I/O simulation
- Enables test automation of diverse software workloads, including unit, integration tests, and fault injection
- Cloud service that can be integrated in CI/CD and **MLOps** development flows



4-5. HPCとエッジAIにおけるNVIDIAとの関係（1/3）

講演：NVIDIA関連のパートナー発表；テクニカルセッションで2件、イベントパートナーセッションで1件、エコシステムセッションで1件、Tech Hubで3件（詳細は次ページ）

(1) 応用される分野、製品

- 全般であるが、結果としてHPC分野とエッジ領域

(2) 当該分野における技術ニーズ

- 高性能コンピューティング、低消費電力、開発期間短縮

(3) 発表（セッション）の概要

- NVIDIAとArmの連携について計7件の発表から現時点での情報を整理

(4) 技術的に特筆すべき点

- 補完関係はサーバーHPCでのArmコアとNVIDIA GPU開発ツールの連携による開発効率化や、エッジ/IoTでのAI/MLソリューションの異なるマシンターゲットへのデプロイの容易性など多岐にわたる。
- 特にHPC分野とエッジ領域ではこの2社の開発ツールの補完関係・連携はかなり強力である。

(5) 本発表（セッション）が注目される理由

- NVIDIAによるARM買収が手続き中であり、全体的に2社でどのようなフォーメーションを目指すのかという観点で発表を眺めた。

(6) 実用化までの課題

- 特になし

4-5. HPCとエッジAIにおけるNVIDIAとの関係（2/3）

- NVIDIA関連のパートナー発表はテクニカルセッションで2件、イベントパートナーセッションで1件、エコシステムセッションで1件、Tech Hubで3件あった（下図）。
- Microsoft, Amazon, Googleなどの発表も多かったのでNVIDIAが特別に多いという訳ではないがNVIDIAによるArm買収手続きが進行中であり、全体的にどのようなフォーメーションを目指すのかという観点で発表を眺めた。
- 全体としてみると、補完関係はサーバーHPCでのArmコアとNVIDIA GPU開発ツールの連携による開発効率化や、エッジ/IoTでのAI/MLソリューションの異なるマシンターゲットへのデプロイの容易性など多岐にわたる。特にHPC分野とエッジ領域ではこの2社の開発ツールの補完関係・連携はかなり強力である。

HPC

クラウドネイティブのスーパーコンピューティング：
ベアメタルのSecured Supercomputing Architecture
Gilad Shainer / NVIDIA

Arm上のCUDA向けHPC開発を簡素化する最新Nsight開発ツール
Jackson Marusarz / Product Manager, Developer Tools / NVIDIA

Armエコシステム向けのNVIDIA HPCソフトウェア開発キット
Jeff Hammond / Principal Engineer / NVIDIA

エッジ

エッジ上のエンドツーエンドのクラウドネイティブ・コンピュータビジョン・アプリケーション
Megh Makwana / Solutions Architect / NVIDIA
Ajeet Raina / Developer Relations Manager / Redis Labs

NVIDIA TritonとArm NNを使用した自動スケーリングのハードウェア非依存型ML推論
Joshua Minor / Research Engineer / Arm

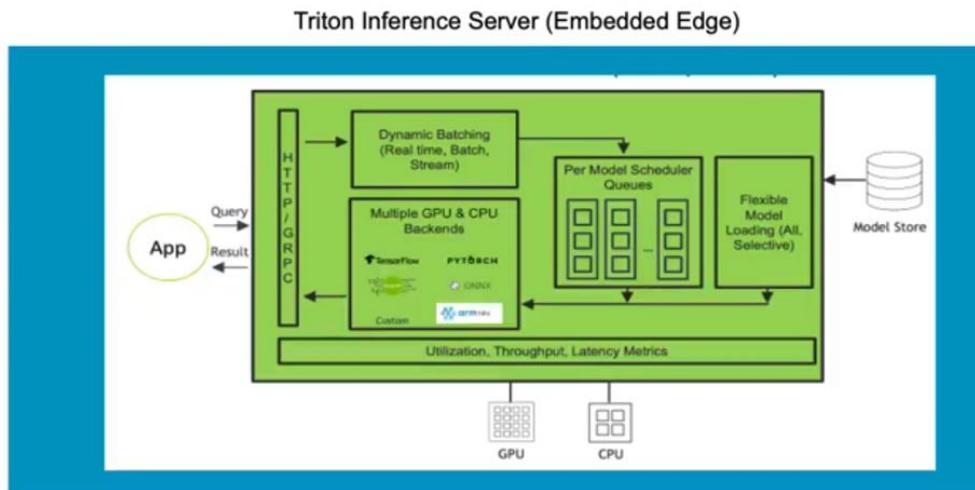
インテリジェントエッジ： Ampere, Arm, NVIDIA, and SUSEを使用したコンピュータ、AI、ソフトウェアエッジ
Bryan Gartner / Senior Technology Strategist / SUSE

エコシステムセッション
神話の崩壊へ： ORAN, COTS, CloudNativeをベースにした超効率的な大規模MiMo 5 Gプラットフォーム
Kuntal Chowdhury / SVP & GM, AI & Analytics, Mavenir
Soma Veyalutham / General Manager, AI and 5G, NVIDIA
Panch Chandrasekran / Director, Carrier Infrastructure Segment Marketing, Arm
Henry Justin Calvert / Senior Director, Head of Future Networks, GSMA

4-5. HPCとエッジAIにおけるNVIDIAとの関係 (3/3)

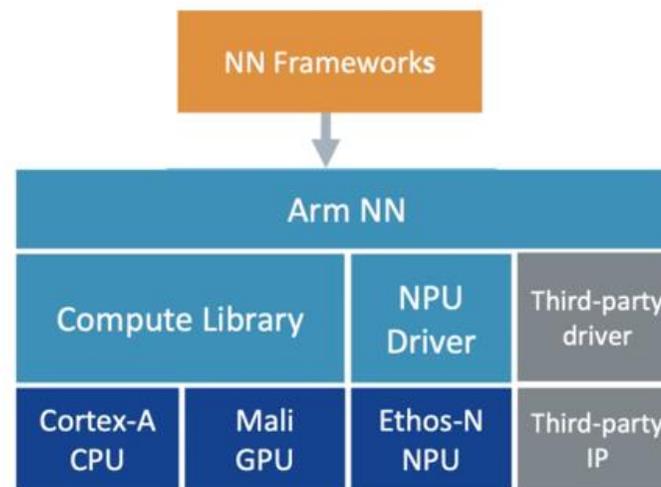
■ Triton/Arm NNによるエッジへのAI/MLデプロイ

- **Triton**はNVIDIAが開発しオープンソースとしたインタフェース・サーバー
- クラウドで開発したモデルをエッジ側のデバイスに移植（デプロイ）する場合、NVIDIAのjetson以外のデバイスにも効率よく展開可能で選択肢が増える
- **Arm NN**はTensorFlow Lite、PyTorch、Android NN API、ONNXなどで開発したニューラルネットワークのモデルをArmのCortex-A CPU、Mali GPU、Ethos-N NPUに最適化するライブラリ
- よく使われるネットワークでTFLiteのままのモデルとArm NNライブラリを使ったモデルの性能を比較すると1.45倍-4.5倍性能向上する



Triton on Raspberry Pi4 (4GB RAM, Arm Cortex A72)

Model Name	TFLite Default + RUY	TFLite + XNNPACK CPU	TFLite + ArmNN CPU
Inceptionv3 – FP32	throughput: 1.4 infer/sec latency: 764 ms	throughput: 1.2 infer/sec latency: 787 ms	throughput: 2.8 infer/sec (2x) latency: 339 ms (2.25x)
Mobilenetv1 1.0 224 – FP32	throughput: 12.4 infer/sec latency: 80 ms	throughput: 10.4 infer/sec latency: 97 ms	throughput: 18 infer/sec (1.45x) latency: 55 ms (1.45x)
Resnet50 v1 – FP32	throughput: 1.8 infer/sec latency: 601 ms	throughput: 1.6 infer/sec latency: 667 ms	throughput: 3.2 infer/sec (1.78x) latency: 314 ms (1.91x)
VGG-16 – FP32	throughput: 0.4 infer/sec latency: 2374 ms	throughput: 0.2 infer/sec latency: 2646 ms	throughput: 1.8 infer/sec (4.5x) latency: 594 ms (4x)
Squeezenetv1.1 – FP32	throughput: 7.8 infer/sec latency: 126 ms	throughput: 8.6 infer/sec latency: 116 ms	throughput: 13.2 infer/sec (1.53x) Latency: 75 ms (1.55x)



4-6. AVCC (Autonomous Vehicle Computing Consortium) (1/2)

講演：Tech Hub 「自動運転車のシステムアーキテクチャの定義」

(1) 応用される分野、製品

- 自動運転

(2) 当該分野における技術ニーズ

- 完全自動運転のサービスを実現しようとする、自動運転技術だけでなく歩行者や周りの車への運転意志表示、乗降の補助、乗員の快適性などの周辺技術も重要であり、複数チップ全体でのトータル消費電力、セキュリティ、開発期間短縮が課題

(3) 発表（セッション）の概要

- AVCC(Autonomous Vehicle Computing Consortium)の紹介、現在の取り組み状況

(4) 技術的に特筆すべき点

- システム定義、情報フロー策定による自動車メーカーと大手サプライヤー、半導体メーカーにまたがった開発の効率化
- 自動運転領域に特化したベンチマークの策定

(5) 本発表（セッション）が注目される理由

- 完全自動運転に向けた業界のコンソーシアムとして標準化団体と独自化・グループ化によるデファクトスタンダードを目指す方向との中間の隙間を埋めるようなタイプのコンソーシアムである。

(6) 実用化までの課題

- 特になし

4-6. AVCC (Autonomous Vehicle Computing Consortium) (2/2)

- 自動車メーカーと大手サプライヤー、半導体メーカーが2019年10月8日、完全自動運転実現のために立ち上げたコンソーシアム

メンバー



AVCC Members Include Organizations from the AV Supply Chain and Ecosystems

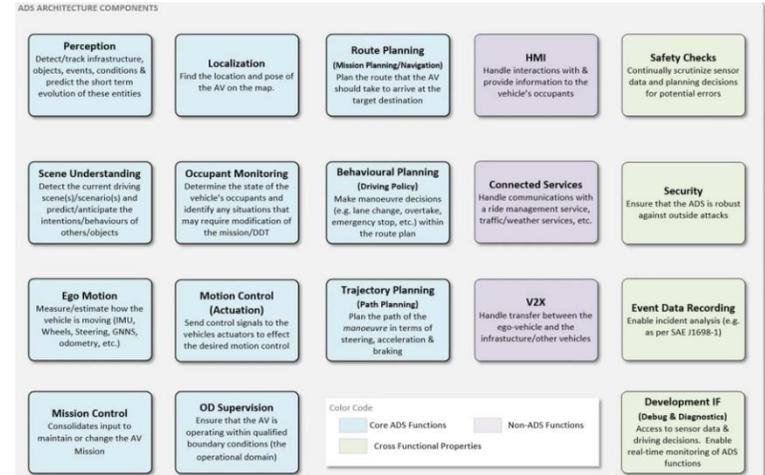
- OEMs
- Tier 1s
- Technology Companies
- Mobility Service Companies
- Commercial Transportation

Entities that inform, govern, influence or help standardize the automobile industry including:

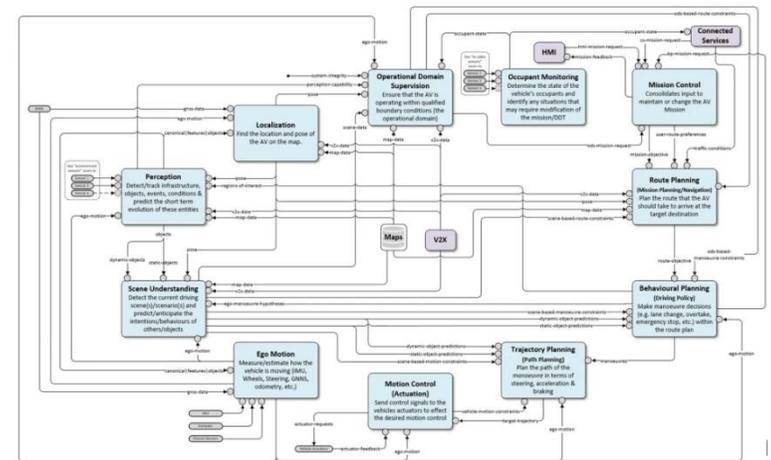
- Government/Regulatory
- Industry Organizations & Groups
- Research & Higher Education

arm DevSummit

システム定義



情報フロー



- 完全自動運転のサービスを実現しようとすると、自動運転技術だけでなく歩行者や周りの車への運転意志表示、乗降の補助、乗員の快適性などの周辺技術も重要になる。
- このためにたくさんのチップが使われるが、消費電力のトータル予算は限られ、複数のチップへの効率の良い機能分散が重要となる。
- 自動運転の競争領域では独自技術を認めるが、非競争領域も大きな部分を占めセキュリティも重要な課題。
- システム定義と情報フローを標準化することにより協力体制を加速する。
- 機械学習でのベンチマークはHPCからエッジ、IoTがMLPerfがカバーしているが、自動運転ではぴったりのものがないので、連携をとりながら自動運転分野でのベンチマーク策定に力を入れている。

2-1-5. NVIDIA GTC 2021 Fall

目次

1. 開催概要
2. NVIDIA GTC 2021 Fall概要
3. 基調講演
 - 3-1. 基調講演概要
 - 3-2. 基調講演紹介技術
4. 基調講演で注目の発表
 - 4-1. OrbNet
 - 4-2. Modulus
 - 4-3. Omniverse
 - 4-4. Hyperion 8

1. 開催概要（1/2）

1. 概要

NVIDIA GTC(GPU Technology Conference)は、2009年から始まったNVIDIAのGPU技術を活用してコンピューティングの課題を解決する事例が多数紹介されるNVIDIAのイベントとして始まり、学会等が主催する国際会議とは位置づけが異なっている。近年は、AIに焦点をあてたグローバルAI会議へとトピックは移り変わっている。

2009年には参加者が1,500人でスタートしたGTCは2021 Springで初めてオンラインに移行し、45,000人の登録者で開催され、オンライン移行後2回目の開催となった2021 Fallでは**21万人**の登録があり、オンライン開催が今後のイベントの開催形式として定着していることが伺える。

NVIDIAはGTCで新技術を毎回発表しており、2021 Springではデータセンタ向けCPUの発表、また、基調講演の舞台となったキッチンが**NVIDIA Omniverse^{*1}**で作成された仮想スペースであったことを後日公表し、話題となった。

2. 開催日時

2021年11月8日 - 11日（4日間）

3. 開催形式

オンラインで開催

*1 NVIDIA Omniverse™は、仮想コラボレーションと物理的に正確なリアルタイム シミュレーションのために開発され、拡張機能に優れたオープン プラットフォーム



URL <https://www.nvidia.com/ja-jp/gtc/>

1. 開催概要（2/2）

4. 構成

講演内容は以下のセッションタイプに属した内容となっており、全体を通じて**522のセッション**が開催された。

- Connect with the Experts: グループまたは1対1でプロジェクトや課題について議論する
- Demo: NVIDIAのテクノロジーをパワフルなデモで紹介する
- DLI Training Lab: インストラクタによるワークショップ
- Keynote: **CEO Jensen Huangによる基調講演 ※NVIDIAの最新情報が披露されるため注目される**
- NVIDIA Inception Startup: GPUを使ったスタートアップの発表
- Special Event
- Talks & Panels

【参考】NVIDIA Corporation 概要

1. 設立

1993年4月5日

2. 本社所在地

米国 カリフォルニア州 サンタクララ

3. マイクロエレクトロニクス業界での位置づけ

4-1. 半導体企業売上高ランキング 7位、Fabless企業で2位

4-2. ディスクリートGPUシェアランキング 1位

(NVIDIA社とAMD社の2社寡占にある)

4. NVIDIAのGPU

GPU(Graphics Processing Units)はその名の通り、グラフィックス処理用のLSIであり、グラフィックス処理に適した演算器が多数搭載されており、並列処理に向けた構造になっている。このGPUの演算能力をグラフィックス処理以外にも使う技術を**GPGPU**(General-Purpose computing on Graphics Processing Units)と呼ぶ。

NVIDIAは早くから(2006年発表)この技術に着目し、**CUDA**(Compute Unified Device Architecture)と呼ぶ開発環境を提供しており、Python等で開発されたプログラムから簡単にGPUの持つ処理能力を利用可能とした。AIブーム、暗号化資産のマイニングブームにより、NVIDIAのGPUが多数使われることとなった。

NVIDIAのGPGPUは、IBMのスーパーコンピュータSummitにも使われている。

5. Arm 買収

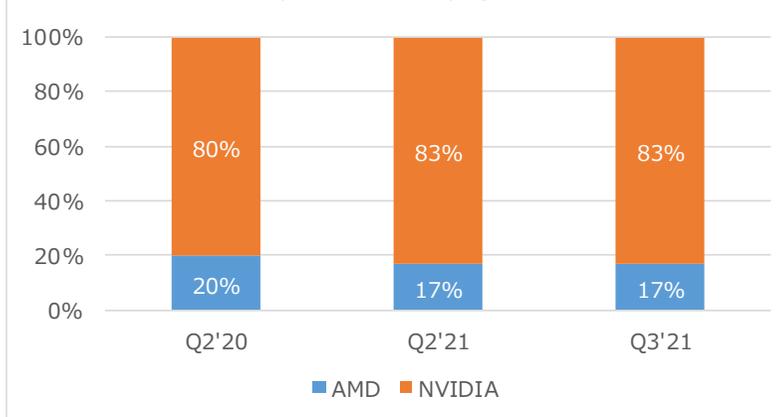
2020年9月14日にArm買収を発表した。しかし、規制当局の承認が得られておらず、買収は完了していない。(2022年1月作成当時)

Semiconductor Companies with >\$10 Billion in Sales in 2021F

2021F Rank	2020 Rank	Company	Headquarters	2020 Total IC	2020 Total O-S-D	2020 Total Semi	2021F Total IC	2021F Total O-S-D	2021F Total Semi	2021/2020 % Change
1	2	Samsung	South Korea	58,555	3,298	61,853	78,850	4,235	83,085	34%
2	1	Intel	U.S.	76,328	0	76,328	75,550	0	75,550	-1%
3	3	TSMC (1)	Taiwan	45,572	0	45,572	56,633	0	56,633	24%
4	4	SK Hynix	South Korea	26,094	981	27,075	35,628	1,639	37,267	38%
5	5	Micron	U.S.	22,542	0	22,542	30,087	0	30,087	33%
6	6	Qualcomm (2)	U.S.	19,357	0	19,357	29,136	0	29,136	51%
7	8	Nvidia (2)	U.S.	14,659	0	14,659	23,026	0	23,026	57%
8	7	Broadcom Inc. (2)	U.S.	15,941	1,803	17,744	18,864	2,099	20,963	18%
9	12	MediaTek (2)	Taiwan	10,985	0	10,985	17,551	0	17,551	60%
10	9	TI	U.S.	12,731	843	13,574	15,889	1,015	16,904	25%
11	15	AMD (2)	U.S.	9,763	0	9,763	16,108	0	16,108	65%
12	11	Infineon	Europe	7,542	3,683	11,225	9,113	4,503	13,616	21%
13	10	Apple* (2)	U.S.	11,440	0	11,440	13,430	0	13,430	17%
14	14	ST	Europe	6,804	3,374	10,178	8,400	4,174	12,574	24%
15	13	Kioxia	Japan	10,553	0	10,553	12,132	0	12,132	15%
16	17	NXP	Europe	7,582	809	8,391	9,711	1,004	10,715	28%
17	19	Analog Devices (3)	U.S.	7,722	405	8,127	9,575	504	10,079	24%
Top-25 Total				364,170	15,196	379,366	459,683	19,173	478,856	26%

(1) Foundry (2) Fabless (3) Includes acquired company's sales in 2020 and 2021 results. Source: Company reports, IC Insights *Custom devices for internal use.

ディスクリートGPU市場シェア



出典：Jon Peddie Researchの情報を基にNTTアドバンステクノロジーが作成

2. NVIDIA GTC 2021 Fall概要 (1/7)

NVIDIA GTC 2021 Fallは、イノベータ、研究者、意思決定者など**21万人**の参加登録があったNVIDIAの年次カンファレンスである。**基調講演の再生数は14百万回**となっている。NVIDIAの最新技術を発表する場であり、NVIDIAのチップを活用した事例がNVIDIAおよびパートナー企業から多数発表される。2021 Fallでは**522のセッション**があり、参加者がセッションに参加しやすいようにセッションで発表されるNVIDIAの技術が活用される業界およびTopicでカテゴライズされており、業界は**18業界**、Topicは**48Topic**と多岐にわたってNVIDIAの技術が活用されているのがわかる。

業界: 18業界

Agriculture, Architecture/Engineering/Construction, Automotive/Transportation, Cloud Service Provider, Consumer Internet, Energy, Financial Services, Game Development, Hardware/Semiconductor, Healthcare & Life Sciences, Higher Education/Research Institution

Topic: 48Topic

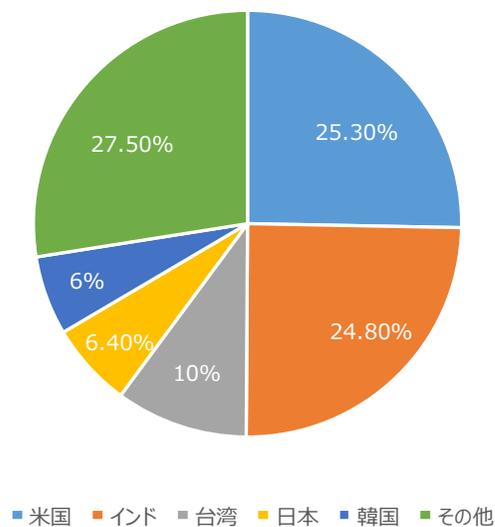
3D Design Collaboration and Simulation, AI Strategy for Business Leaders, Accelerated Computing & Dev Tools - Algorithms/Numerical Techniques, Accelerated Computing & Dev Tools - Libraries/Runtimes, Accelerated Computing & Dev Tools - Performance Optimization, Accelerated Computing & Dev Tools - Profilers/Debuggers/Code Analysis, Accelerated Computing & Dev Tools - Programming Languages/Compilers, Autonomous Machines - Reinforcement Learning, Autonomous Machines - Robotics, Autonomous Machines - Robotics Research, Autonomous Vehicles, Building & Infrastructure - Design/Engineering/Construction, Computer Vision - Research, Conversational AI/NLP, Cybersecurity/Fraud Detection, Data Center - Networking, Data Center - Virtualization, Data Center/Cloud - Business Strategy, Data Center/Cloud Infrastructure - Technical, Data Center CPUs, Data Science, Deep Learning - Frameworks, Deep Learning - Inference, Deep Learning - Training, Graphics - AI Applications/Art, Graphics - Animation/VFX/Virtual Production, Graphics - Production Rendering and Ray Tracing, Graphics - Real-Time Rendering and Ray Tracing, HPC - Astronomy/Astrophysics, HPC - Climate/Weather/Ocean Modeling, HPC - Computational Chemistry and Materials Science, HPC - Computational Fluid Dynamics, HPC - Computational Physics, HPC - Quantum Computing, HPC - Scientific Visualization, HPC - Supercomputing, Healthcare - Drug Discovery/Genomics, Healthcare - Medical Imaging, Healthcare - Smart Hospitals & Instruments, IoT/5G/Edge, Manufacturing - Inspection/Predictive Maintenance/Logistics, Product Development - Design/Engineering and Manufacturing, Recommenders/Personalization, Retail - Intelligent Stores/Logistics/Data Science, Video Streaming/Conferencing, XR(Virtual and Augmented Reality)

2. NVIDIA GTC 2021 Fall概要 (2/7)

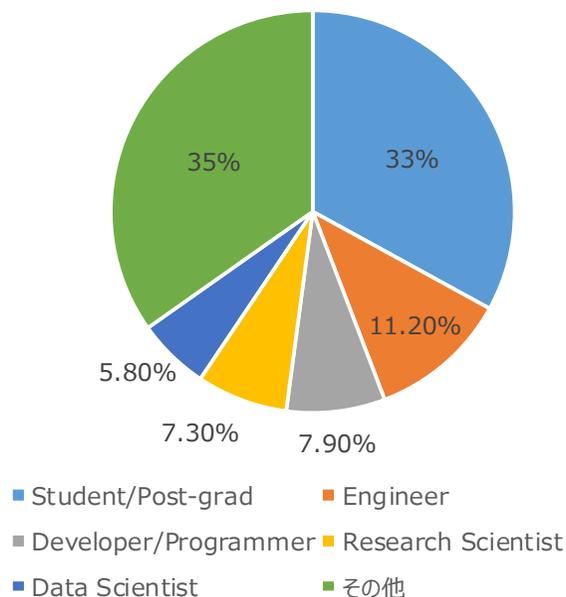
参加者属性

参加者の25.3%が米国からの参加で最も多く、次がインドからの参加者で24.8%であった。
参加者の職務別割合では、学生が33%と最も多く、次が11.2%でエンジニアであった。
参加者の業界別割合では、高等教育・研究が最も多く23.6%で次がIT Serviceであった。

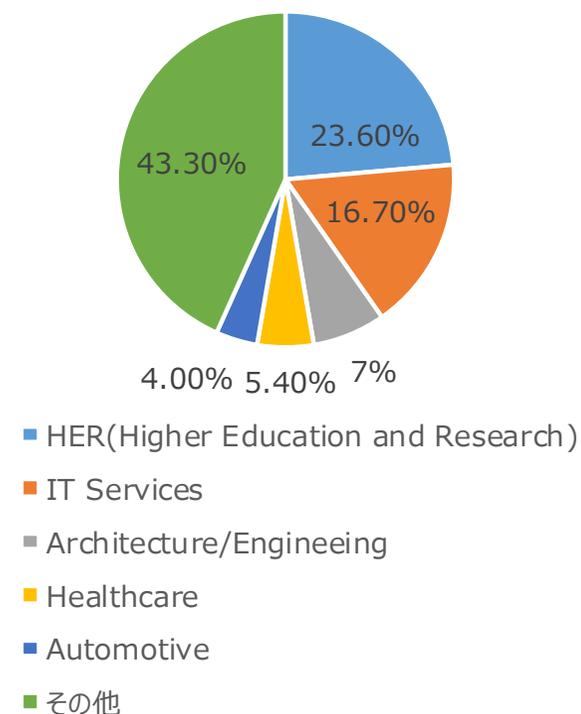
参加者の参加場所別割合



参加者の職務別割合



参加者の業界別割合



(資料) <https://www.nvidia.com/gtc/sponsors/> の情報を基にNTT-ATが作成

2. NVIDIA GTC 2021 Fall概要 (3/7)

全体を通して以下の17セッションは日本語での開催もあった。

セッション番号 / Topic	タイトル	概要	発表者 / 所属
A31580/ IoT, 5G, Edge	5G・MECによるクラウドレンダリングサービスの実証	5GとAWS Wavelengthを基盤としたクラウドレンダリングの技術やシステム構成について、これまで実施した複数のイベントにおける実証実験の内容（デモ動画）を交えながら解説する。特に4G・クラウドとの品質差分、および、AR スマートグラス・バーチャルヒューマン・デジタルツインなどを用いることによる体験価値を示し、新しいネットワーク環境やXRアセットを利用する意義について明らかにする。	福井啓充/KDDI
A31562/ Autonomous Machines – Robotics	DeepStream SDK マルチカメラへの応用	世界中に存在する数十億台のカメラは大量のデータを取り込んで、ビジネスインサイトの創出、プロセスの効率化、収益源の改善に活用できる。AI ベースのマルチセンサー処理やビデオ、音声、画像の理解のためのストリーミング分析ツールキット、DeepStream SDK の特徴として複数の映像ソースの取り込みによるパイプラインの構築が可能。本セッションでは、CSI カメラ/USB カメラを用いた複数映像ソースのリアルタイムでのDeepStreamパイプラインの構築について解説する。	佐々木陽/FaBo
A31615/ Deep Learning – Frameworks	“DL for DX” DXにおけるディープラーニング導入動向について	DX時代、あらゆるビジネスシーンにおいてAI・ディープラーニング活用を視野に入れた変革が必要になっている。その変革に、企業や団体、またビジネスパーソンはそれぞれどのように向かっていけば良いのか。人材育成やデジタルリテラシー整備の観点から、日本におけるAI・ディープラーニング活用の最新動向や事例なども交えて話す。	岡田隆太郎/日本ディープラーニング協会
A31613/ Deep Learning – Frameworks	DXに成功する企業とそうでない企業の差は？ 武蔵精密工業の大塚社長が語る「組織像・経営者像」にせまる	武蔵精密は、AI技術と製造に関する深い知識と経験を組み合わせることでDXを可能にした日本の業界のトップランナーの1つである。NVIDIA AI DAYSでの前回のセッションに続くこのセッションでは、CEOの大塚宏が成功への道と他の産業のブレークスルーにまたがる最新の成果を深く掘り下げる。	井崎武士/NVIDIA 大塚浩史/武蔵精密工業

(資料) <https://events.rainfocus.com/widget/nvidia/nvidiagtc/sessioncatalog?search.language=1594320459782002LzDW>の情報を基にNTT-ATが作成

2. NVIDIA GTC 2021 Fall概要 (4/7)

セッション番号 / Topic	タイトル	概要	発表者 / 所属
A31614/ Deep Learning - Frameworks	DXの思考法	<ul style="list-style-type: none"> ・DXについて経営者が理解しなければいけないのは、「基本的な視座」であり「思考法」である。 ・デジタル化で破壊的なイノベーションを起こした人に共通の発想は、「分野に関係なく一つのやり方で問題を一気に解決できるのではないか」という抽象化の発想である。 ・DXとは「コーポレート・トランスフォーメーション」(CX)であり、組織風土を含めた経営改革だが、同時に、ソフトウェアなどがもつ基本的な性格を理解し経営をそれに合わせる必要がある。 ・デジタル化は経済社会を縦割りからレイヤー構造に転換するので、DXは必然的に産業丸ごとの転換、つまり「インダストリアル・トランスフォーメーション」(IX)になる。 ・DXを進めるには、「本棚にない本を探す」という発想、つまりすでにあるデジタルツールはそれを借りてきて使いこなし、ないものに注力して開発し、他社に提供する、という発想が必要である。 	西山圭太/経営共創基盤, 東京大学未来ビジョン研究センター
A31628/ Data Center - Networking	IoT/AI時代に求められる次世代データセンターテクノロジー	近年、IoTやAIの需要が大幅に高まる中、データセンターに求められる要件も大きく変遷し、時代に合わせた変革を求めている。本講演では、これからの時代に求められる、IoTやAIの需要、利用に応えられる、次世代のデータセンターに求められる技術や要件をひもとくとともに、これを実現するNVIDIAのソリューションを紹介する。	大西宏之/NVIDIA
A31558/ Data Center - Virtualization	NVIDIA vGPUでDXを推進！ NVIDIAが提案する次世代仮想基盤	多くの企業は仮想化技術を取り入れ、業務システムの統合を進めてコスト削減や運用性などのメリットを得てきた。守りのシステムであった仮想化技術は進化を続け、従来あきらめていたシステム、そして新たな利用用途への可能性を広げている。パワフルな最新Ampere世代GPUと柔軟性をもたらすNVIDIA vGPU(仮想GPU)、そして仮想基盤の効率を大きく向上させるDPUを組み合わせ、DX推進仮想基盤として、あらゆるワークロード、VDIやVR、レンダリングやデザインコラボレーション、AI、機械学習、HPC(科学技術計算)と言った、性能要件や専門性の高いニーズにまで応えらえる、次世代の仮想基盤を紹介する。	後藤祐一郎/NVIDIA
A31565/ Graphics - Real-Time Rendering and Ray Tracing	Omniverse Enterprise製品紹介、ライセンスと必要な構成	Omniverse Enterprise の導入に向けて、製品版でのコンポーネント構成とライセンス、システム構築に必要なRTX、ワークステーション、サーバーの構成を紹介する。	So Takahashi/NVIDIA

(資料) <https://events.rainfocus.com/widget/nvidia/nvidiagtc/sessioncatalog?search.language=1594320459782002LzDW>の情報を基にNTT-ATが作成

2. NVIDIA GTC 2021 Fall概要 (5/7)

セッション番号/Topic	タイトル	概要	発表者/所属
SE31564/ Computer Vision – Image Processing	Omniverse User Group Japan: ここが凄い！Omniverseで変わる3Dコンテンツ制作のワークフロー	Omniverseで何がかわるか、3Dコンテンツ制作でのメリットは何か？Omniverseを先行して試しているクリエイター、エンジニアの方をお招きして、ズバリ語っていただきます。Omniverse調べているけど、どこに聞けばよい？強力なパートナーチームをご紹介します。このセッションはパネルディスカッションのライブ配信。	Hideaki Tanaka等 /NVIDIA, Prometech, Autodesk, ASK, Quebico, GDEP Advance, Too
A31627/ Conversational AI/NLP	NVIDIA RIVAを用いてあなたの対話AIを構築する	NVIDIA Riva SDKについて紹介する。ASR、NLU、TTSモジュール、および多言語（英語、日本語）の事前トレーニング済みモデルを含み、マルチモーダル会話型AI指向のデモも紹介する。車の運転支援、Omniverseとの接続による3Dキャラクター構築など、いくつかの実際のアプリケーションシナリオでRivaを適用するデモもまじえ、エンドツーエンドのASR（citriNETなど）およびTTSモデル（tacotron2、waveglow、flowtronなど）も紹介する。	Xianchao Wu/NVIDIA
A31626/ Healthcare – Medical Imaging	カスケード型医工連携と連合学習	画像診断を含む保健医療分野でのAI開発で最も障害になっている要因の一つに個人情報保護の課題がある。この課題の克服方法のうち今年注目されているトピックの一つに連合学習がある。東京慈恵会医科大学人工知能医学研究部では、この連合学習の手法と従来のAIの学習結果を比較する研究を行っているため、その一端を紹介するとともに連合学習におけるエッジコンピューティングの重要性について解説する。	中田典生/東京慈恵会医科大
A31568/ Accelerated Computing & Dev Tools - Profilers / Debuggers / Code Analysis	デザインレビューに向けたOmniverse Enterprise構成方法	建築デザイン、CG制作でのレビューを想定し、Omniverse Enterpriseのインストールからセットアップに必要な手順を、技術的な視点から解説する。	Sammy Kakizawa/NVIDIA
A31586/ Data Center / Cloud Infrastructure - Technical	ノーコード時系列データ分析ツールNode-AIの紹介と、その基盤を担うKubernetesベースのGPUプラットフォーム	NTTComでは機械学習プロジェクトにおけるコミュニケーションを効率化することを目的としたノーコードで時系列分析が可能なツール、Node-AIを内製開発している。今回はNode-AIの概要と、それを社内で支えるGPUおよびKubernetesベースのプラットフォームについて解説する。	切通恵介, 紀本雅大 /NTTコミュニケーションズ

(資料) <https://events.rainfocus.com/widget/nvidia/nvidiagtc/sessioncatalog?search.language=1594320459782002LzDW>の情報を基にNTT-ATが作成

2. NVIDIA GTC 2021 Fall概要 (6/7)

セッション番号/Topic	タイトル	概要	発表者/所属
A31561/ Data Center - Networking	パフォーマンスを妥協することなく次世代のゼロトラストのセキュリティ実現する	セキュリティに関する要求は高まる一方である。しかし多くの人々は、要求の全体像であったり、ボトルネックとなるポイントを理解することに苦労しており、結果としてセキュリティ要求を満たせていなかったり、性能、スケーラビリティに問題があり、ユーザ体験に問題があるシステムを設計してしまう。セキュリティとパフォーマンスやスケーラビリティはトレードオフの関係にあるべきではない。このセッションでは昨今のセキュリティ要求のトレンドから、それを満たすために我々はどのように変革すべきなのか、そしてDPU（データプロセッシングユニット）を含むNVIDIA製品、技術を利用してこの難題に応える道筋を解説する。	愛甲浩史/NVIDIA
A31601/ Conversational AI / NLP	マルチモーダル環境における Personalized AIサービスを支えるAPI基盤のご紹介	QAコミュニティサービス「教えてgoo」における恋愛相談AIオシエルや、複数の著名なTVドラマキャラクタのチャットボットの開発などで、日本の対話サービスをリードしてきた。対話型AIは広く普及し、現在では、様々なサービスをユーザが享受する際のユーザインタフェースとして不可欠なものとなりつつある。本講演では、NTTレゾナントが進める、音声、映像、言語の3つのモダリティを統合し、対話、レコメンド、サービス代行、メンタルウェルネス、エンタテインメントチャット、ユーザマッチングなど、幅広いサービスを支える基盤となるAIアルゴリズムAPI基盤について紹介する。本API基盤を用いることで、クライアント企業様の希望するサービスを、スピーディに構築することができる。今回、本基盤を支える、Personalized AI技術やマルチモーダル技術を簡潔に紹介し、実現例として、「ウィットに富んだ雑談を音声対話AIと交わりながらユーザ代行でフードデリバリーサービスを手配するデモ」と、「映像、音声、言語を統合しユーザとAIの感情や会話を学習し、ユーザのアバタとAIキャラクタがエンタテインメント会話をを行うデモ」の2つを紹介する。	中辻真, 立石修平, 小瀬木悠佳/NTTレゾナント
A31559/ HPC - Climate / Weather / Ocean Modeling	植物フェノミクスのためのイメージセンシングとAI	植物フェノミクスは、植物の成長、性能、組成を学習する新たな学際的研究分野である。機械学習、コンピュータビジョン、リモートセンシング、ロボティクスなどの学際的な技術を統合して、植物の機能や環境との相互作用を表す形質を表現する。本講演では、ドローンリモートセンシング、3Dイメージング、限られたラベル付きデータセットを用いた深層学習、エッジコンピューティングなど、私たちの研究室で行われているいくつかの研究テーマについて紹介する。更に、このような高度なAI関連技術を用いて、ハイスループット植物フェノタイプング技術開発の加速、農業への応用を紹介する。	Wei GUO/東京大学

(資料) <https://events.rainfocus.com/widget/nvidia/nvidiagtc/sessioncatalog?search.language=1594320459782002LzDW>の情報を基にNTT-ATが作成

2. NVIDIA GTC 2021 Fall概要 (7/7)

セッション番号/Topic	タイトル	概要	発表者/所属
A31585/ IoT, 5G, Edge	5G×MEC×GPUで実現される世界 ～ 5Gならではのユースケース ～	【パート1】VR・AI・モビリティなどの低遅延ソリューションへのニーズが高まる中、私たちはMECサービス「ドコモオープンイノベーションクラウド」(ドコモOIC)を提供してきた。これまでドコモOIC上に構築した様々なソリューションのいくつかを紹介する。5G×MEC×GPUの世界の可能性を感じてほしい。 【パート2】MEC (ドコモイノベーションクラウド) 上で、XRリモートレンダリングサービスを開発した。GPUを搭載した MECを活用することで、性能の低いVRやMRの端末向けに、ハイクオリティなコンテンツのリモートレンダリングを可能とした。SaaS形式でのサービス提供を予定しているので、ユーザはVR/MRコンテンツを用意するだけで、誰でもかんたんに利用できる。その仕組みや利用方法、事例などについて紹介する。	Yoshikazu Akinaga, 西田卓爾/NTTドコモ

(資料) <https://events.rainfocus.com/widget/nvidia/nvidiagtc/sessioncatalog?search.language=1594320459782002LzDW>の情報を基にNTT-ATが作成

3. 基調講演

3-1. 基調講演概要 (1/2)

NVIDIA GTC 2021 Fallの基調講演もSpringと同じOmniverseで作成されたキッチンからスタートした。

講演冒頭で、NVIDIAの技術を活用したAI事例を以下の7つの「i am」で紹介した。

「i am an explorer」

- ・北極海での安全な航路の探索
- ・月面の水分探索

「i am a visionary」

- ・気候変動シミュレーション
- ・CO2地下貯蔵予測

「i am a healer」

- ・入院患者の見守り支援
- ・治療方針の策定支援

「i am a storyteller」

- ・多言語発話
- ・作曲

「i am a guardian」

- ・衛星から地上の汚染、山火事を監視

「i am a helper」

- ・農業支援

「i am a creator」

- ・製造現場の支援
- ・新しい表現手法の創出

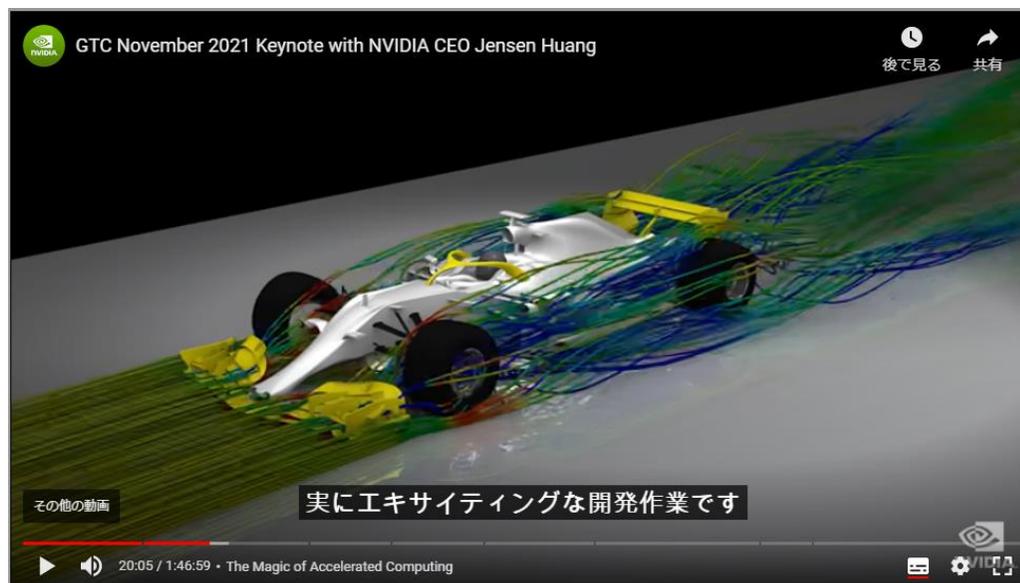


NVIDIA CEO Jensen Huangの基調講演
https://www.youtube.com/watch?v=jhDiaUL_RaM

3-1. 基調講演概要 (2/2)

【NVIDIAの実績】

- GPUから始まり、科学・産業アプリケーションの各分野でハードウェアからソフトウェアまでのフルスタックを作成しており、ゲーム・デザイン・生命科学・地球科学・量子コンピューティング・AIサイバーセキュリティ・5G・ロボティクスにまでわたる**150種以上のSDK(Software Development Kit:ソフトウェア開発に必要なプログラム、ライブラリ、サンプル等をパッケージ化したもの)**を用意している。
- NVIDIAを使用する開発者の数は**過去5年間でほぼ6倍の300万人に、CUDA^{*2}は過去15年間で3,000万回、昨年だけで700万回ダウンロード**されている。
- GTC 2021 Fallでは**65の新規及び更新されたSDK**が紹介された。



*2 NVIDIAのGPU用汎用並列コンピューティングプラットフォームおよびプログラミングモデル。専用のC/C++コンパイラ やライブラリ などが提供されている。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

3-2. 基調講演紹介技術（1/7）

基調講演で紹介された技術を以下に示す。※技術名が赤字の技術は後述している。

#	技術名	技術概要
1	ReOpt	配達車両のルーティングや倉庫のピッキングと梱包などオペレーションズ・リサーチの最適化問題を高速で解決するソルバー。 超並列アルゴリズムは、何千ものソリューション候補と改良を生成する。 動的な再ルーティングにより、移動時間が短縮され、10兆ドルのロジスティクス業界で数十億ドルを節約できる。
2	cuQuantum	GPU上で実行する 量子回路シミュレーションを高速化するSDK 。アプライアンス型のディープラーニング用スーパーコンピュータDGX上で、状態ベクトルとTensorネットワーク法を用いて量子回路シミュレーションを高速化し、1,688量子ビットを使用して3,375個の頂点の最大カットのソリューションを発見する量子アルゴリズムシミュレーションの大きなマイルストーンを達成した。
3	cuNumeric	Pythonの数値演算ライブラリNumPyアクセラレータライブラリ。 コード変更なしにNumPyを高速化し、GPUからマルチGPU、マルチノードクラスタ、さらには世界最大のスーパーコンピュータにスケーリングする。 NumPyは過去5年間で1億2,200万回ダウンロードされており、GitHubの約80万件のプロジェクトで使用されている。cuNumericは、cuDF・cuML・cuGraphを含むNVIDIAのRAPIDSオープンソースPythonデータサイエンススイートの一部である。NVIDIAの Legion 上の構築されている。 データセンタ規模の数値演算ライブラリ。 スケーラビリティは優秀で、流体力学の基礎を学習するためのCDF Pythonティーチングコードを使った例では わずか20%の損失で1,000基のGPUにスケーリングしている。
4	Legion	タスクをCPU・GPU・DPUにスケジュールする データセンタ規模のコンピューティングエンジン 。タスクレベルの並列処理を抽出し、これらのタスクの実行をデータセンタ全体に動的に並べ替えて処理する。
5	Toy-Me	対話型アバター 。音声理解、自然言語処理、自身の音声による音声合成、キャラクターアニメーション、美しいレイトレーシンググラフィックス、これら 全てをリアルタイムに実行 する。基調講演ではCEOのアバターが登場し、質疑応答するデモ映像が流れた。
6	Quantum-2	400Gbps のInfiniBandプラットフォーム。InfiniBandとは、非常に高いRAS（信頼性・可用性・保守性）を持つ基幹系・HPC系のサーバ/クラスター用高速I/Oバスアーキテクチャ及びインターコネクトのこと。システム間インターコネクト機構としては、RAS機能の他、他機構に比較して、低レイテンシである点も特徴である。 Quantum-2スイッチ、ConnectX-7NIC、BlueField-3DPU 等で構成される。



(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

3-2. 基調講演紹介技術 (2/7)

#	技術名	技術概要
7	ConnectX-7	これまでもっとも高速なNIC(Network Interface Card)。400 Gbpsのラインレートでの暗号化が可能。16コア、256スレッドのデータパスプロセッサを搭載。RDMA(Remote Direct Memory Access)、GPU-Direct Storage、GPU-Direct RDMA、in-networkコンピューティングがConnectX-6の2倍になる。
8	BlueField-3	インフラストラクチャコンピューティングプラットフォーム。データセンタインフラストラクチャをオフロードして分離するための64bit Arm CPUコアを16基搭載。ネットワークストレージおよびセキュリティソフトウェアの処理にかかるCPU負荷を軽減するためのDPU(Data Processing Unit)。インフラストラクチャソフトウェアは約30%ものCPUを消費している。
9	BlueField DOCA 1.2	サイバーセキュリティ機能スイート。ゼロトラストセキュリティプラットフォーム構築のためのプラットフォームである。境界やワークグループセグメントでの保護はもはや不十分である。アプリケーション、データ、ユーザ、デバイスのすべてのタッチポイントが潜在的な攻撃対象になる。BlueFieldはネットワークのエンドポイントとなるため、ほぼすべてのタッチポイントでデータセンタを保護できる。BlueFieldのエコシステムは拡大しており、Palo Alto Networks、Checkpoint、Juniper、Fortinet、Guardicore、Trend Micro、F5、VMWareなどの大手サイバーセキュリティ企業がBlueFieldで次世代ファイアウォールサービスを提供するためNVIDIAと協力している。
10	Morpheus	すべてのユーザ、マシン、サービスによって生成されたわずかなデータセンタの特性を監視し、分析できるディープラーニングサイバーセキュリティプラットフォーム。NVIDIA RAPIDSとNVIDIA AIに基づいて構築されている。Morpheusのワークフローは、アプリとユーザのあらゆる組み合わせに対してAIモデルデジタル指紋を作成し、通常のパターンを学習して異常なトランザクションを探索する。MorpheusはGPUコンピューティングのパワーを活かし、環境にあわせてモデルをカスタマイズし、以前は不可能であった方法でネットワークを保護する。
11	PINO	物理法則に従うユニバーサル関数学習機。物理法則に従うようトレーニングされた物理モデルであるPINN(Physics-Informed Neural Networks)と偏微分方程式の近似を学習できるFNO(Fourier Neural Operator)の利点を組み合わせて構築されており、物理的原理に基づいたシミュレータまたは観測データから学習できる。トレーニングが完了すると物理的原理に基づいたモデルを高速にシミュレートできる。もうひとつ重要な点として、このモデルは並列性が高いため非常に大規模なシステムにスケールさせることで100万倍の速度向上を得ることができる。

3-2. 基調講演紹介技術 (3/7)

#	技術名	技術概要
12	OrbNet	機械学習とコンピューティングを使用して創薬に革命を起こすことを目指しているEntosが開発した 物理機械学習(物理ML) の手法による グラフニューラルネットワーク 。EntosはNVIDIA Clara Discoveryを活用して研究を進めており、基礎をなす量子力学に関する性質をより多く機械学習モデルに採用している。候補となる医薬品はタンパク質のアミノ酸との化学結合を形成している。これらの反応はまれにしか起こらないので長期間シミュレーションする必要がある。静電場シミュレーションでは原子間結合をモデル化できないため、反応の自由エネルギーを計算するには量子計算が必要である。OrbNetがなければ 3ヶ月以上かかるシミュレーションを1つのGPUで3時間で実行する 。
13	Modulus	物理MLモデルを開発するためのフレームワーク。現象を支配する物理と原理モデルおよび観測により得られるデータを使用して物理MLモデルを学習する。Modulusは マルチGPUとマルチノードでの学習に最適化 されている。 作成されるモデルはシミュレーションよりも1,000-10万倍速く物理的現象をエミュレーション できる。科学者はこれまでになく大きなシステムの理解を深めるためにModulusで デジタルツインを作成 することができる。
14	Omniverse	仮想コラボレーションと物理的に正確なリアルタイム シミュレーション のために開発され、拡張機能に優れたオープン プラットフォーム。Omniverseは デザインの世界をつなぐ 。たとえば、Adobeの世界で作られたものとAutodeskで作られたものをつなぐことができ デザイナーは共有空間で共同作業 ができる。ある世界のデザイナーによる変更は3Dデザインのクラウド共有ドキュメントのようにつながっているすべてのデザイナーに対して反映される。また、企業はOmniverseで 仮想工場を構築し仮想ロボットを稼働 させることができる。仮想工場と仮想ロボットは物理的レプリカの デジタルツイン である。物理バージョンはデジタルから生成されるためデジタルのレプリカである。Omniverseのデジタルツインは 未来のロボット、ビル、工場、倉庫、車を設計・学習 および継続的に監視する場所である。 2020年末の提供開始以来500社のデザイナーによって7万回ダウンロードされている。
15	Omniverse Showroom	Omniverseの コアテクノロジー(グラフィクス、物理学、マテリアル、AI) を紹介するための デモやサンプルアプリを収録 している。
16	Omniverse Farm	複数のシステム(ワークステーション、サーバ、ベアメタルまたは仮想環境)の間での バッチジョブの処理を調整するシステムレイヤ 。バッチレンダリングやAIの合成データ生成、分散コンピューティングに使用。
17	Omniverse AR	携帯電話やARグラスにグラフィクスをストリーミングする。

3-2. 基調講演紹介技術 (4/7)

#	技術名	技術概要
18	Omniverse VR	世界初のフルフレームインタラクティブレイトレーシングVR。
19	iTwin	Bentleyのデジタルツイン構築基盤。大規模な産業および土木インフラストラクチャプロジェクトのデジタルツインをフォトリアスティックかつリアルタイムに視覚化およびシミュレーションするため、 NVIDIA Omniverse プラットフォームを使用した。エンジニアリング企業の90%が利用しており、Bentley iTwinのユーザー数は 約200万人 になる。
20	Siemens Energy	NVIDIA Modulusを用いて多層乱流用のデジタルツインを開発している。点群データを用いて物理ベースのAIモデルをトレーニングすることで流線で示される忠実度のフローを数秒で推論できる。基調講演では、排熱回収ボイラー(HRSG)の腐食予測の事例を紹介した。 Siemens Energy の推定によれば腐食を正確に予測することで定期メンテナンスでの検査と計画外の ダウンタイムを70%削減 できる。業界全体で5~6日短縮すると 年間20億ドル近く節約 できる。HRSGの腐食は複数の物理的問題で、流動特性、水化学、動作条件が組み合わされる。
21	Isaac Gym	強化学習 研究のためのNVIDIAの 物理シミュレーション環境 。Isaac GymはOmniverseで構築されている。基調講演では、 BMW が自動車組立工場に組立ロボットに新しいスキルを学習させている事例を紹介した。BMWの継続的な改善とイノベーションの要求に対応するためにはアウトプットを早め、俊敏性を高め、どこでも効率を最適化するための複雑な生産ナリオをシミュレーションする必要があり、OmniverseはBMWが重要な意思決定を行い、自動化を導入する必要があるすべての場所で新しいシミュレーション機能を導入できる。
22	Omniverse Extension	Omniverseの拡張機能。基調講演ではEricssonが開発した、電波の伝搬データを統合し、OmniverseのRTX高速レイトレーシングを活用し、 都市のあらゆるポイントで信号の品質を速やかに可視化および計算できるようにしたExtension を紹介した。OmniverseでEricssonは 建物、植生、葉のマテリアルまで物理的に正確な都市規模のモデルを構築 する。そして無線ネットワークコンポーネントが追加され、各トランスミッターの正確な場所、高さアンテナパターンが指定される。 Omniverseマテリアルは物理的に正確であるため、反射の強度が正確に決定される 。アンテナビームフォーミングと信号経路を正確にシミュレーションし、可視化できる。ビジュアライゼーションはEricssonにとって重要な機能である。

3-2. 基調講演紹介技術 (5/7)

#	技術名	技術概要
23	Nemo Megatron	数十億あるいは1兆におよぶパラメータをもつ音声モデルと言語モデルのトレーニング専用のフレームワーク。巨大なシステムにも対応し、最高の計算効率を維持できるように最適化されている。NVIDIAの研究者はNVIDIAの500ノードSelen DGX SuperPODでGPT-3を11日間でトレーニングした。5,300億のパラメータがあるMegatron MT-NLGのモデルはマイクロソフトと共に6週間でトレーニングした。
24	TensorRTコンパイラ	推論ソフトウェア開発のための環境。リアルタイム推論を可能とする。最適化とデプロイの2段階でソフトウェアを作成する。2020 FallではTensorFlowとPyTorchにネイティブに統合されたことを発表した。多くの開発者がこれらのフレームワークから直接推論しており、簡単で確実に動作するが遅いのが欠点であった。TensorFlowにこれらのフレームワークが統合されたことにより、これからは1行のコードでML開発者は指一本動かすことなく推論を3倍速くできる。
25	Triton推論サーバ	NVIDIAのGPUを使用して機械学習モデルを高速に推論させるサーバを構築するためのフレームワーク。これまでのMLモデルに加えDLでの推論にも対応することを発表した。推論プラットフォームが1つあれば、TritonのGPUとCPUでDLとMLを推論できるようになった。あらゆるモデル、あらゆるフレームワーク、複数のクエリ型で推論する。MLとDL、あらゆるプラットフォーム、クラウド、オンプレミス、エッジ、組み込み、にマルチGPU、マルチノード、CUDA、x86、Armで、あらゆる推論ワークロードを1つのエンジンNVIDIA Tritonで画像処理から音声認識、音声合成、自然言語処理、レコメンダー、強化学習までTritonのパフォーマンスはあらゆる面で優れている。
26	UCF(Unified Computing Framework)	ロボティクスパイプラインを処理するAIアプリケーション作成のためのフレームワーク。専用アクセラレータ、CUDA GPU、Tensor コアAI、RTXグラフィックス、ネットワーク、セキュリティ、高速I/Oの処理を連携させることでコンテナやマイクロサービスを高速パイプラインに構成できる。UCFではデータセンターまたは組み込みシステムで実行されるアプリケーションを構築できる。
27	Metroplis	ビデオ処理および解析プラットフォーム。ストリーミングビデオから検出、追跡、カウント、3Dポーズ推論が可能で、さらに将来的には完全な3Dシーン再構築に対応する。現在カメラをサポートしているが、UCFで簡単に拡張可能でLiDAR、深度センサ、画像処理レーダー、超音波、赤外線に対応する。

3-2. 基調講演紹介技術 (6/7)

#	技術名	技術概要
28	MAVedge-AI	ソフトウェアデファインド5Gソリューションで業界をリードするMavenir社がMetropolisプラットフォームを利用して開発した5G無線ネットワーク構築キット。工場、プラント、公共スペース、農場、そしてITが制限される場所で産業用アプリケーションのAI-on-5Gを実現する。2022年第一四半期に先行アクセスとして提供予定。
29	Magnum IO	データセンター向けのIOサブシステムであり、マルチテナントデータセンターをサポートするIOと通信を高速化するための新しい拡張機能を導入するもの。2020 FallではVerizonとの共同研究事例として、Magnum IOを使用し、ビデオストリームを10倍の速さで書き込み30分あたり最大250TBのビデオを処理する事例を紹介した。
30	Maxine	アバタープラットフォーム、つまり仮想ロボットプラットフォームである。Maxineは自律型や遠隔操作型、リアルの用途やアートの用途で利用できる。Maxineはカスタマーサービスなどの広範囲にわたるアプリケーションで利用できる。現実世界やWebまたOmniverse内など形態もさまざまである。ビデオ会議でも利用できる。また、ゲームのキャラクターをアニメーション化したりロボットに統合することもできる。
31	Riva	ニューラルスピーチAI。これは、Maxineの入力と出力になる。Rivaは英語を話し、7つの言語(英語、スペイン語、ドイツ語、フランス語、日本語、北京語、ロシア語)を認識する。将来はもっと多くの言語を話す予定である。Rivaは字幕化、翻訳、要約、質問への回答、意図の理解ができる。Rivaの認識精度は世界最高レベルであり、応答時間は他の追従を許さない。そしてわずか30分のトレーニングでRivaは特定の声に、たとえば企業ブランドアンバサダーの声に合わせることができる。Rivaは声の高さ、長さ、感情に合わせて微調整され、人間のように表現する。Rivaはあらゆるクラウドやエッジに展開できる。
32	Tokkio	NVIDIAが会話型キオスクと呼ぶMetropolisのアプリケーション。この小さなアニメーションロボットはお客様の目を見て動きを目で追う。音声認識から、意図や行動を推論する自然言語のダイアログマネージャ、レコメンデーション、自然な会話までTokkioは約2秒以内で応答し、しかもとてもインタラクティブである。



(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

3-2. 基調講演紹介技術 (7/7)

#	技術名	技術概要
33	Clara Holoscan	ソフトウェアデファインドで プログラミング可能な画像処理プラットフォーム 。未来の医療機器は ロボット を使用することになり、この産業は未来を構築するためにソフトウェアデファインドの画像処理プラットフォームを必要としている。HoloscanはNVIDIAテクノロジーのすべてを導入して作られている。UCFと、超高速センサー処理のロボティクスチップである Orinチップ が使われている。Holoscan UCFを導入することでこうしたパワフルなエンジンを接続する リアルタイムアプリケーションを作るための開発プラットフォーム が機器メーカーに提供される。Holoscanはオープンプラットフォームで、 サードパーティはHoloscanのインタフェースとAPIをベースに開発 できる。研究者は優れた新しい科学を実践し、 機器メーカーはHoloscanを自社のソリューションに統合 できる。HoloscanはNVIDIAの 第3のロボティクスプラットフォーム である。
34	Orinチップ	超高速センサー処理ロボティクスチップ。センサ、物理演算、AI画像処理、グラフィックスからなる ロボティクスパイプライン全体を1つのチップで処理 できる。12個のArm CPU、 5.2TFLOPS のFP32、 250TOPS のAI性能、毎秒740Gbの高速I/Oでセンサを接続、オプションで A6000 Ampere GPUを追加 すればFP32が39TFLOPS増え、 AI性能が600TOPSを超える 。
35	Isaac Sim Replicator	ロボットをトレーニングするための合成データを生成するエンジン。 センサをシミュレーションし、自動的にラベルが付けられたデータを生成し、ドメインランダム化エンジンを使用して豊富で多様なトレーニングデータセットを作成 する。路上の自動車とは違い、ロボットの世界は遥かに無作為であるため、ロボティクスではトレーニングデータの作成は非常に困難である。
36	DRIVE	フルスタックの オープンAV(Auto Vehicle)プラットフォーム 。NVIDIAの顧客はNVIDIAの開発フローの利用、NVIDIAの自動運転用コンピュータの利用、NVIDIAのクラウドマップへの接続、もしくはNVIDIAとエンドツーエンドでパートナーシップを結ぶことができる。 自動運転車両はロボット であり、機械学習の開発と同じ3本の柱がある。 NVIDIA AIを使ったモデルのトレーニング、Omniverse上のDRIVE Simによるシミュレーションや合成データの生成、Orinロボティクスチップ上のDRIVE AVによるリアルタイムロボティクスパイプラインの3本柱 である。
37	Hyperion 8	自動運転車プラットフォーム 。完全に運用可能で、量産車の開発を前提としたオープンな自動運転車プラットフォームで、AI 機能や自動運転に必要なテクノロジーを車両に搭載するのに必要な時間とコストを大幅に削減する。センサスイートは12台のカメラ、9台のレーダー、12台の超音波センサ、1台の前方LiDARで構成される。これがすべて 2基のOrinで処理 される。開発者向けキットではAmpere GPUを追加することでその性能に余裕を持たせており、エンジニアは最良の環境で新たなソフトウェアを試作できるようになる。Hyperion には、 最高レベルの性能の自律運転システムを検証するのに必要なハードウェアがすべて含まれている 。

4. 基調講演で注目の発表

- 基調講演で紹介され、ニュース等で注目された技術および協業成果等の技術を次スライド以降で紹介する。当該技術のベースとなる技術、ベースとした技術についても紹介する。

#	技術名	技術概要
12	OrbNet	Entosとの協業成果の発表。物理MLによりタンパク質のアミノ酸との化学結合をシミュレーション。
13	Modulus	Caltech(カリフォルニア工科大学)との協業成果の発表。物理現象を高速にエミュレーションし、デジタルツインを作成する。
14	Omniverse	仮想コラボレーションと物理的に正確なリアルタイムシミュレーションのオープンプラットフォーム。GTC 2021 Springの基調講演でも話題になった。Bentley、Siemens、BMW、Ericssonとの協業成果の発表。
37	Hyperion 8	自動運転車プラットフォーム。自動運転を実現するためにNVIDIAの様々な技術をベースにしている。基調講演ではNVIDIAの技術を使った車内環境の進歩についても紹介された。

4-1. OrbNet (1/2)

(1) 本発表が応用される分野、製品

化学物質とタンパク質の間の分子間力を分子動力学的にシミュレーションして候補物質を探索する創薬分野

(2) 当該分野における技術ニーズ

膨大な量の分子シミュレーションの高速化

(3) 本発表の概要

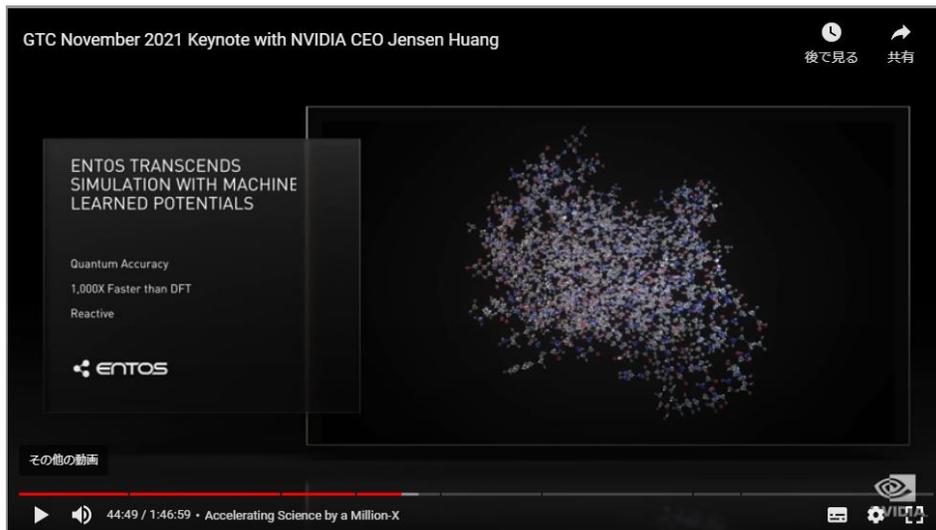
化学物質とタンパク質の間の分子間力を分子動力学的にシミュレーションする処理は現代の創薬の重要な要素になっている。分子の原子間力はその3D構造によって決定される。ヒトタンパク質の3D構造アミノ酸の長鎖はX線結晶構造解析とクライオ電子顕微鏡により明らかにされる。約25,000個のヒトタンパク質のうち、この高度な方法を使用して解読されたタンパク質はわずか17%。人間の病気は膨大な種類があり、ヒトタンパク質の3D構造がわからなければコンピュータを活用した創薬の適用範囲も限定されたものになる。AIモデルは既知の有効な化学物質の特性を学習して効果が見込まれる他の新しい化学物質を生成できるようになった。効果が見込まれる何百万もの化学物質と数十万ものタンパク質構造を組み合わせることで新しいチャンスが膨大に広がる。これにより、今度は膨大な量の分子シミュレーションがボトルネックになる。

基調講演では、EntosのOrbNetを使って他のタンパク質の適切な折り畳みを促進するシャペロンタンパク質であるHSP-90*³のシミュレーションの事例を紹介した。これは、HSP-90タンパク質と候補薬品の間で起こっている化学反応のシミュレーションである。候補となる医薬品はタンパク質のアミノ酸との化学結合を形成している。これらの反応はまれにしか起こらないので長期間シミュレーションする必要がある。静電場シミュレーションでは原子間結合をモデル化できないため、反応の自由エネルギーを計算するには量子計算が必要である。このシミュレーションは1つのGPUで3時間かかった。

物理MLであるOrbNetがなければ3ヶ月以上かかる。

*3 他のタンパク質が適切に折りたたまれ、熱ストレスに対してタンパク質を安定化させ、タンパク質の分解を助けるシャペロンタンパク質。また、腫瘍の成長に必要な多くのタンパク質を安定化させるため、抗がん剤としてHSP-90阻害剤が研究されている。

4-1. OrbNet (2/2)



(4)技術的に特筆すべき点

物理機械学習(ML)の手法により、分子シミュレーションにおける原子間力の高コストな量子計算のかわりとしてグラフニューラルネットワークを学習させた。

(5)本発表が注目される理由

OrbNetを使うことにより、3ヶ月かかる計算を3時間に短縮できた事例の発表である。

(6)実用化までの課題

既にEntosで使われており、特になし。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-2. Modulus (1/3)

(1) 本発表が応用される分野、製品

気候予測等の物理的現象をエミュレーションする**デジタルツイン***4分野。

(2) 当該分野における技術ニーズ

エミュレーションの高速化

(3) 本発表の概要

物理MLモデルを開発するためのフレームワーク。現象を支配する物理と原理モデルおよび観測により得られるデータを使用して物理MLモデルを学習する。ModulusはマルチGPUとマルチノードでの学習に最適化されている。作成されるモデルはシミュレーションよりも**1,000-10万倍速く物理的現象をエミュレーションできる**。科学者はこれまでになく大きなシステムの理解を深めるためにModulusで**デジタルツイン**を作成することができる。

Modulusで解決できる重要な問題のひとつに気候科学がある。気候変動は世界を変えつつある。アクセラレーテッドコンピューティング、物理MLおよび巨大なコンピュータシステムを組み合わせると100万倍の飛躍を実現し、実行することができる。NVIDIAでは物理的原理モデルと観測データを使用してAIに気候を瞬時に予測するように学習させる。継続的に未来を予測しながら、そのデータを元にモデルを調整して改善することで**地球のデジタルツイン**を作ることが可能になる。

Caltech(カリフォルニア工科大学)とNVIDIAの研究者はECMWF(欧州中期予報センタ)のERA5の大気データを使用して物理MLモデルを学習した。このモデルを128 A100 GPUで学習すると4時間かかる。学習が完了すると、この物理MLモデルはハリケーンの重大度とその経路を予測できるようになる。GPUでの7日分の予測は1/4秒しかかからない。**これはシミュレーションの10万倍の速さ**である。

*4 物理空間から取得した情報をもとに、デジタル空間に物理空間の双子（コピー）を再現する技術。工場や製造設備の建設、都市開発など、あらゆる現場においてデジタル空間に物理空間を再現することによって、事前のシミュレーション・分析・最適化を行い、それを物理空間にフィードバックする仕組み全体を指す。

4-2. Modulus (2/3)

GTC November 2021 Keynote with NVIDIA CEO Jensen Huang

ANNOUNCING NVIDIA MODULUS
Physics-ML Neural Simulation Framework

Framework for Developing Physics-ML Models
Train Physics-ML Models Using Governing Physics, Simulation, and Observed Data
Multi-GPU, Multi-Node Training
1,000-100,000X Speed Models - Ideal for Digital Twins

Available Now
developer.nvidia.com/modulus

その他の動画

47:39 / 1:46:59 • Accelerating Science by a Million-X

後で見る 共有

The diagram illustrates the Modulus framework architecture. It starts with a 'SymPy Equation' (e.g., $\frac{dy}{dx} = \tau \cdot (y^2) - \theta$) which is processed by a 'Computational Graph Compiler' (using Model Library like ISIRI, PINN, PINN, MESHFREE). This leads to a 'Multi-Node Multi-GPU Training Engine' which integrates 'Geometry ICs & BCs', 'Observations', and 'Numerical Optimization Plans'. The engine is shown running on a multi-processor hardware architecture.

GTC November 2021 Keynote with NVIDIA CEO Jensen Huang

2016, Sep 17

Surface Winds [mph]

25 35 45 55 65

Actual Track

48:55 / 1:46:59 • Accelerating Science by a Million-X

後で見る 共有

The video shows a globe with weather prediction overlays. A legend indicates 'FNO Prediction' (purple circles) and 'Ground Truth' (pink circles). A 'NOAA Forecast' track is also shown. The video title is 'Deep Neural Operator Physics-ML Model Predicts Extreme Weather'.

(4)技術的に特筆すべき点

物理MLエンジンにNVIDIAのGPUアクセラレーション用cuDNN、NVIDIA MagnumIOを利用して高速化を図っている。

(5)本発表が注目される理由

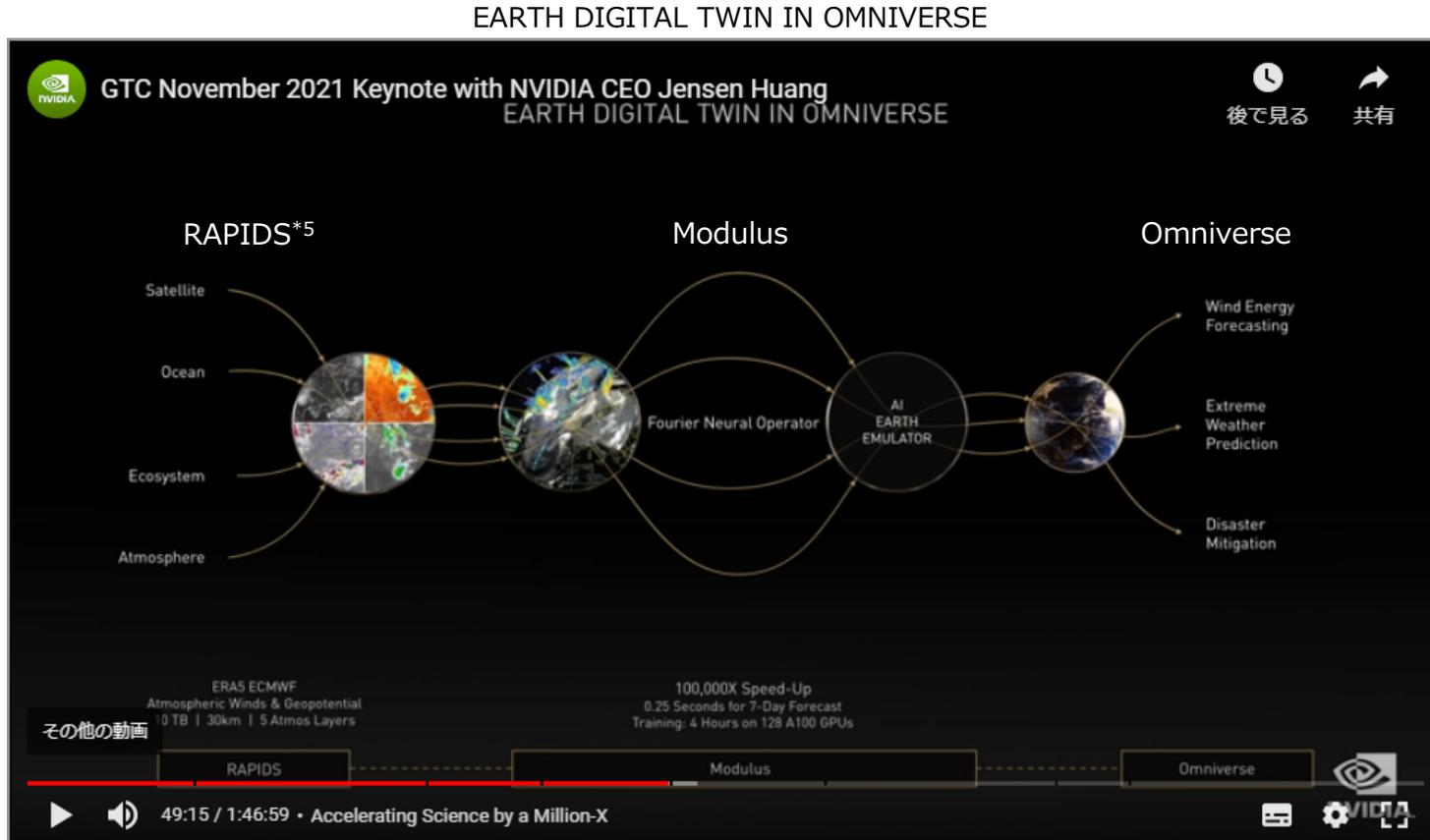
Modulusで作成されるモデルはシミュレーションよりも1,000-10万倍速く物理的現象をエミュレーションできる。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-2. Modulus (3/3)

(6) 実用化までの課題

基調講演では、数年後にはOmniverseで実行されている地球のデジタルツインにデータがストリーミングされ、物理MLモデルを組み合わせることで気候を予測できるようになるだろうと紹介していた。



*5 データサイエンスのワークフロー全体をGPUで高速化するためのライブラリ群。GPUの性能を引き出すNVIDIA CUDAベースで構築され、使いやすいPythonインタフェースを提供している。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-3. Omniverse (1/4)

(1) 本発表が応用される分野、製品

アーキテクチャ・エンジニアリング・建設業界、メディア・エンターテインメント業界、製造業界、サイエンス、ゲーム業界等多岐にわたる。

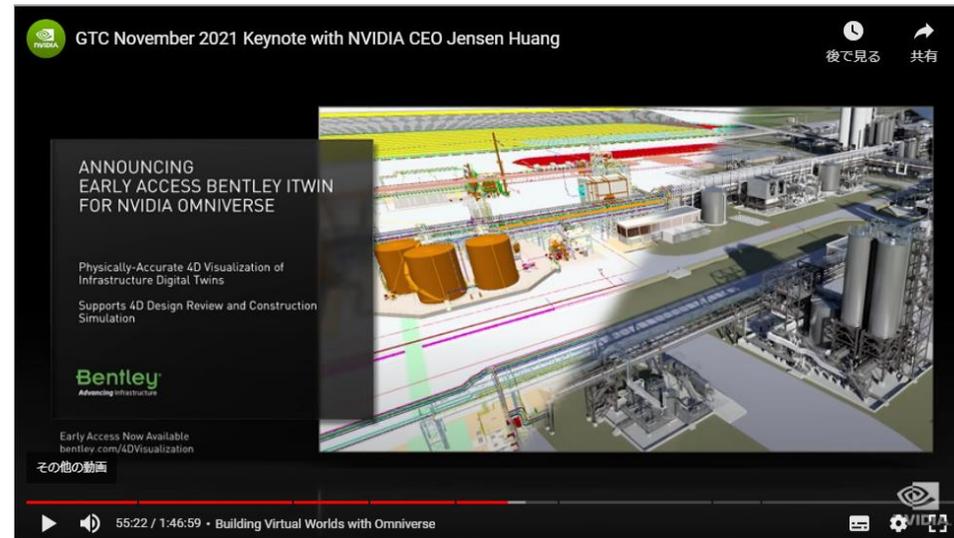
(2) 当該分野における技術ニーズ

ビジュアライゼーション、正確なシミュレーション、短期間でのレビューと発見のために物理的に正確で美しいレイ トレース レンダリング

(3) 本発表の概要

仮想コラボレーションと物理的に正確なリアルタイムシミュレーションのために開発され、拡張機能に優れたオープンプラットフォームの紹介。2020年末の提供開始以来Omniverseは500社のデザイナーによって7万回ダウンロードされている。基調講演では、Omniverseを利用したBentley、Siemens、BMW、Ericssonの事例を紹介した。

【Bentley iTwin】
インフラデジタルツインの
ためのプラットフォーム



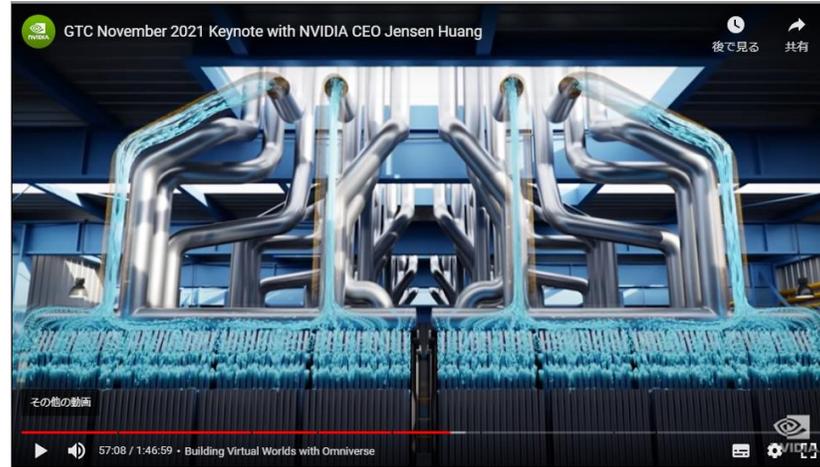
BentleyはOmniverseを採用したiTwinの早期アクセスを発表した。Omniverseに接続するだけでなく、Omniverseで独自のデジタルプラットフォームも構築している。エンジニアリング企業の90%が利用しており、Bentley iTwinのユーザ数は約200万人。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-3. Omniverse (2/4)

【Siemens Energy】

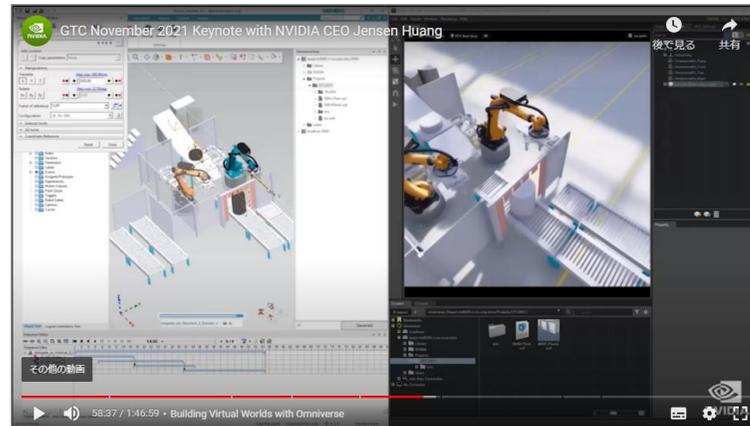
排熱回収ボイラーの腐食を予測する
デジタルツインプラットフォーム



Modulusを用いて多層乱流用のデジタルツインを開発している。点群データを用いて物理ベースのAIモデルをトレーニングすることで流線で示される忠実度のフローを数秒で推論できる。これにより**ダウンタイムが70%削減**され、**年間17億ドルのコスト削減が期待**できる。

【BMW「未来の工場」】

Omniverseで構築されたIsaac Gymを使用してロボットに新しいスキルを学習させている



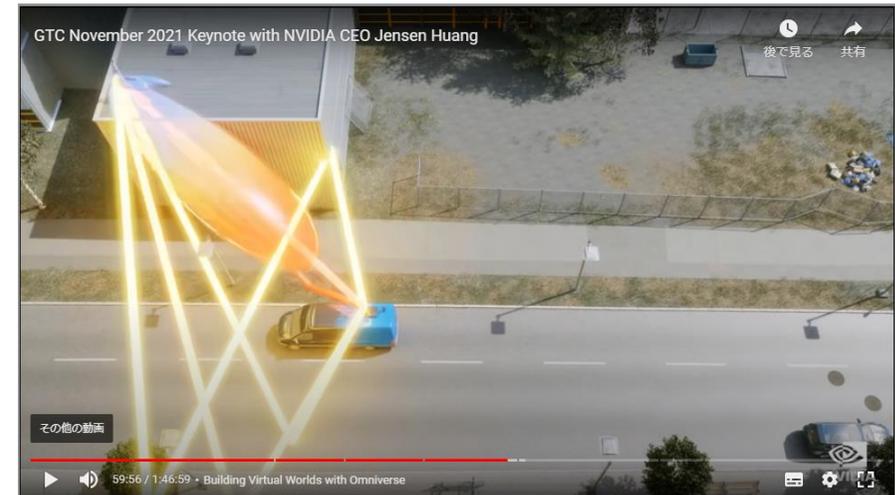
継続的な改善とイノベーションの要求に対応するためにはアウトプットを早め、俊敏性を高め、どこでも効率を最適化するための複雑な生産シナリオをシミュレーションする必要がある。NVIDIA OmniverseはBMWが重要な意思決定を行い、**自動化を導入する必要があるすべての場所で新しいシミュレーション機能を導入可能とする。**

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-3. Omniverse (3/4)

【Ericsson Ominiverse Extension】

電波の伝搬データを統合し、OmniverseのRTX高速レイトレーシングを活用し、都市のあらゆるポイントで信号の品質を速やかに可視化および計算できるようにした。

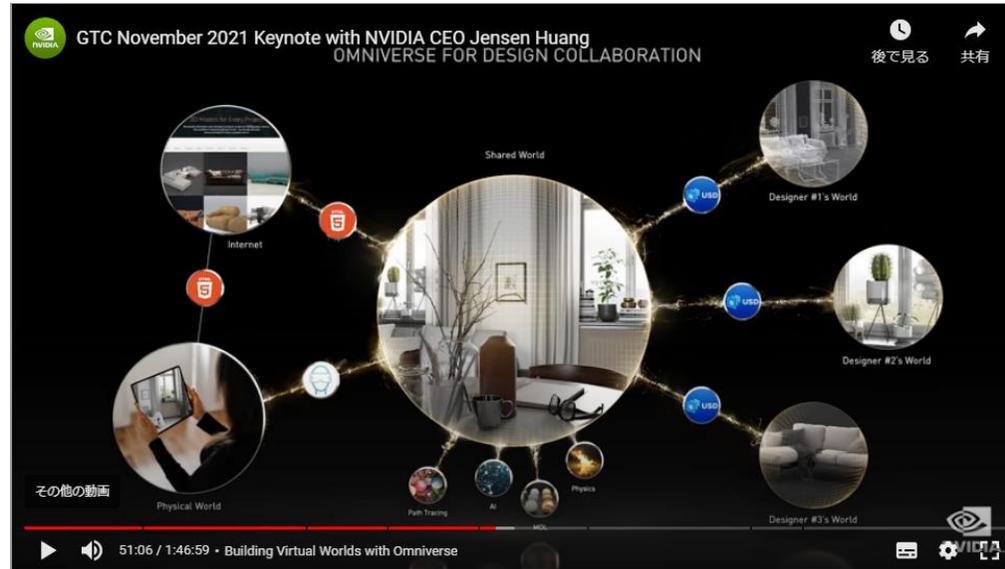


Ericssonは都市全体のデジタルツインを構築し、自社の一連の5Gアンテナと無線機の構成、運用および継続的な最適化を行っている。今後5年間で1,500万の5Gマイクロセルとタワーが世界中で展開される予定になっている。Ericssonはデジタルツイン環境の構築にOmniverseを使用し最高のカバレッジとネットワークパフォーマンスのために各拠点をどのように配置し、構成するかを決定するのに役立っている。OminiverseでEricssonは建物、植生、葉の材料まで物理的に正確な都市規模のモデルを構築する。そして無線ネットワークコンポーネントが追加されると、各トランスミッターの正確な場所、高さ、およびアンテナパターンが計算される。

ネットワークエンジニアは事実上1:1スケールで世界中のあらゆる場所にレポートしモデルのあらゆる部分を探索できる。**ネットワークを調整してパフォーマンスを最適化するときや経路の妨害を特定するとき、調整の効果を文字通りリアルタイムで「見る」ことができる。**OmniverseではEricssonが5Gネットワーク全体を現実に忠実にリモートシミュレーションでき今までより効率的で信頼性が高いネットワークの設計、リモートフィールドのトライアルの実施、展開の迅速化が可能になる。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-3. Omniverse (4/4)



(4)技術的に特筆すべき点

Omniverseは仮想コラボレーションのためにNVIDIAの様々な技術をベースに使い、リアルタイムシミュレーションを実現している。

(5)本発表が注目される理由

FacebookがMetaに社名変更するなどメタバース（コンピュータやコンピュータネットワークの中に構築された現実世界とは異なる3次元の仮想空間やそのサービス）が世界的に注目されているなかでのGPUトップ企業NVIDIAの仮想コラボレーションプラットフォームの発表であるため。ちなみに、GTC 2021 Springでは、基調講演のキッチンがOmniverseで作成されたものであることが、後日発表され、誰も仮想的に作成されたものとわからないほど精巧であったことで注目を集めた。

(6)実用化までの課題

既に様々な企業で使われており、特になし。

(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-4. Hyperion 8 (1/5)

(1) 本発表が応用される分野、製品

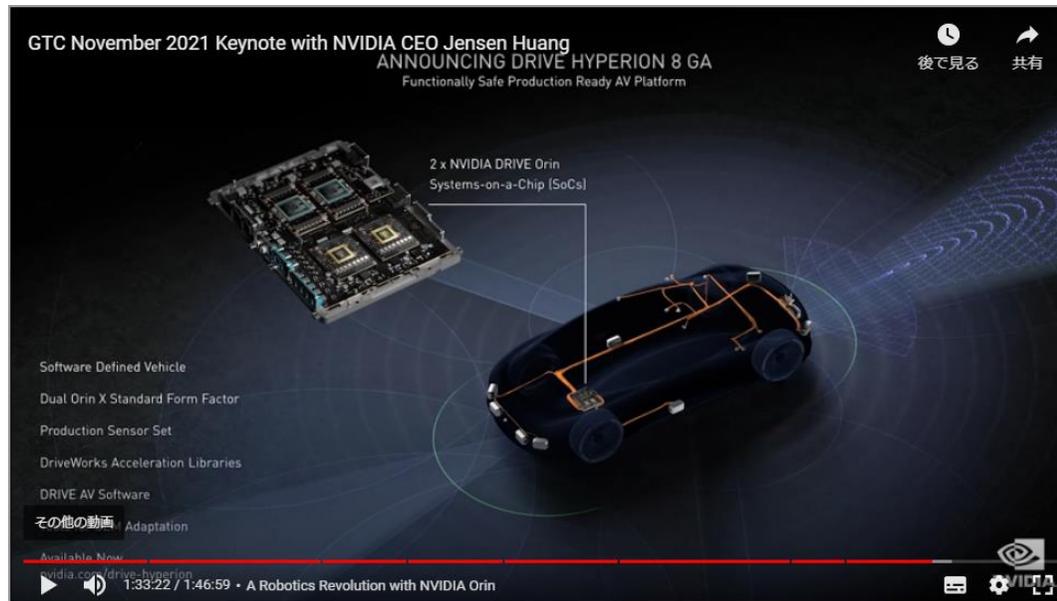
自動運転車

(2) 当該分野における技術ニーズ

- ・自動運転に必要なリアルタイム処理に対応したハードウェア・ソフトウェアを総合的にテストする技術
- ・日々の過酷な運転条件に対する耐久力

(3) 本発表の概要

自動運転車プラットフォームであるHyperion 8は、センサスイートとして12台のカメラ、9台のレーダー、12台の超音波センサ、1台の前方LiDARで構成される。これがすべて2基のOrin^{*後述}で処理される。開発者向けキットではAmpere GPUを追加することでその性能に余裕を持たせており、エンジニアは最良の環境で新たなソフトウェアを試作できるようになる。AVやAV向けセンサの開発者にとってHyperion 8は最適なプラットフォームである。



(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-4. Hyperion 8 (2/5)

NVIDIAは世界中からペタバイト単位の道路データを収集し、研修を受けた約3,000人のラベラーがトレーニングデータを作成している。NVIDIAデータ戦略の土台になるのはやはり合成データ生成である。



下のスライドにシミュレーションされたサ라운드カメラによるシーンがある。データに自動的にラベルが付けられる。右側のAIモデルの一部はこれらのデータでトレーニングされている。DRIVE Sim Replicatorは自動運転車両のための合成データジェネレータであり、Omniverseをベースに開発されている。レンズモデルをシミュレートし、モーションブラー、ローリングシャッター、LEDフリッカー、ドップラー効果を考慮する。

NVIDIAはセンサメーカーと緊密に連携し、センサを正確にモデル化している。カメラ、レーダー、LiDARのセンサモデルはパストレース(レンダリングの一種)で情報を生成する。Omniverseの中の素材は物理的にシミュレートされ、光源からの光を正確に反射する。NVIDIAはLiDARマテリアルライブラリを構築し、レーダーマテリアルライブラリも構築中である。

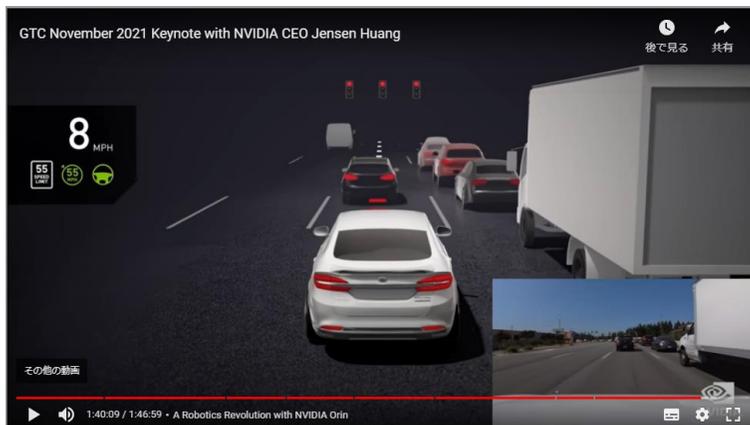


(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-4. Hyperion 8 (3/5)

NVIDIAはGTC 2021 Springの後、都市部での自動運転と自動駐車の手組みを始めた。現在Hyperion 8は、センサ4D認識DLベースのマルチセンサフュージョン、特徴追跡、新しいプランニングエンジンを実行している。AVは車の運転方法を一変させ、道路の安全性を大きく向上させる。

車内にも一大進歩がもたらされる。Maxineのテクノロジーは人と車のあり方を新たにする。Maxineで車はコンシェルジュになる。Maxineはお抱え運転手のように振る舞い、ニューラルグラフィックスを使用して3Dサラウンドビューを再構築し利用者に提示する。



駐車の際は、駐車代行モードによって自動で場所を探し駐車する。さらに、MaxineはNVIDIAが開発しているその驚異的なアバターテクノロジーにより素晴らしいスマートさを発揮する。このすべては新しいOrinロボティクスチップで実行される。

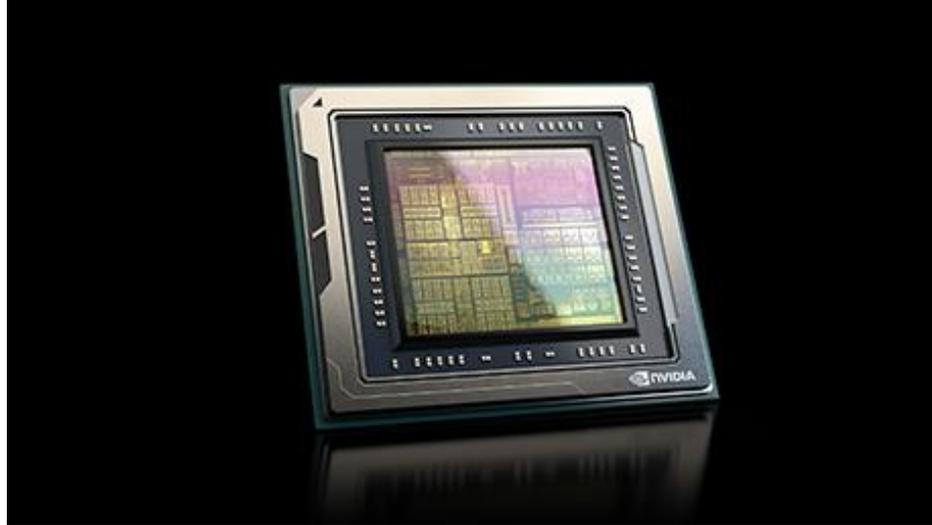


(資料) https://www.youtube.com/watch?v=jhDiaUL_RaMのスクリーンショットをNTT-ATが作成

4-4. Hyperion 8 (4/5)

【Orinチップ】

Orinチップはセンサ、物理演算、AI画像処理、グラフィックスからなるロボティクスパイプライン全体を1つのチップで処理できる。12個のArm CPU、5.2TFLOPSのFP32、250TOPSのAI性能、毎秒740Gbの高速I/Oでセンサを接続、オプションでA6000 Ampere GPUを追加すればFP32が39TFLOPS増え、AI性能が600TOPSを超える。



(資料) <https://www.nvidia.com/ja-jp/self-driving-cars/drive-platform/hardware/>

4-4. Hyperion 8 (5/5)

(4)技術的に特筆すべき点

単なるマスコットとしてのデジタルエージェントではなく、車両の制御を担うシステムとの統合を基に自動運転中の車両が周囲をどのように認識し、どう行動しようとしているかを、低遅延に高精度なCGで示す。AIプラットフォーム間の連携により、駐車した車両の呼び出しに対応し、駐車スペースの探索などについても利用者に情報を伝えることができる。

(5)本発表が注目される理由

- ・自動車大手のメルセデス・ベンツと連携した走行テストであること。
- ・自動運転車に使われているNVIDIAの様々な技術が実環境でうまく働くことを示している点。

(6)実用化までの課題

実証不足が課題。今後さらに実証を重ねる必要がある。

2-1-6. SC21

目次

1. 開催概要
2. SC21概要
3. Technical Paper Award
4. キーノート、招待講演、受賞講演
5. SCおよび特筆すべき発表の概要
 - 5-1. 特筆すべき発表の概要
 - 5-1-1. TOP500 BOF
 - 5-1-2. Green500 BOF
 - 5-1-3. Revealing Power, Energy and Thermal Dynamics of a 200PF Pre-Exascale Supercomputer
 - 5-1-4. Closing the “Quantum Supremacy” Gap: Archiving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputer

1. 開催概要

1. 開催日時

2021年11月14日 – 19日（6日間）

日程	内容
11/14-15,19	Workshop/tutorial
11/16-18	Keynote, Paper, Poster, BoF, etc.
11/15-18	Exhibition

2. 開催形式

ハイブリッド： St. LouisとHUBB(リモート)

3. 主催

ACM, IEEE Computer Society

4. SC概要

この国際会議は、1988年にDOE関係の国立研究所や大学関係者によっていわばボランティア的に企画され、ACMとIEEE/CSが共同主催し、年を追って規模が拡大してきた。テクニカルな論文発表や基調講演・招待講演を主とする一般的なカンファレンスと、企業並びに研究機関の大規模な展示会が併設されているのが特徴である。また、BOF(Birds of Feather)と呼ぶ特定のテーマの自由討論会も昼休みや夕方に多数開催され、そこで**TOP500**などのスーパーコンピュータのランキングなどが発表されることでも注目されている。また、重要な科学技術計算における最大性能を実現したことに対する**Gordon Bell賞***について、ファイナリストによる論文発表と表彰が行われる。

* ACMが毎年授与する、高性能コンピュータ分野で顕著な業績のあったシステムや技術を讃える賞。



URL <https://sc21.supercomputing.org/>

2. SC21概要

ACMとIEEE Computer Societyが主催する、スーパーコンピューティングに関する世界最大の国際会議SC21: The International Conference for High Performance Computing, Networking, Storage and Analysis (通称SC21またはSupercomputing 2021) は、34回目の今年、2021年11月14日(日)から19日(金)までミズーリ州St. LouisとHUBBを用いたリモートのハイブリッドで開催された。テーマは“**science & beyond.**”である。

Technical Programは以下のように様々なセッションが設けられた。投稿数、採択数についてはPapersとResearch Postersにおいて情報が開示された。

セッション	発表数	投稿数	採択数	採択率
Papers	98	361	99	27%
Workshops	38	-	-	-
Tutorial	36	-	-	-
Panels	25	-	-	-
Birds of a Feather(BOF) Sessions	70	-	-	-
Research Posters	43	71	45	63%
ACM Student Research Competitions	23	-	-	-
Doctorial Showcase	10	-	-	-
Scientific Visualization Showcase	6	-	-	-

※BOFセッションからTOP500 BOF、Green500 BOFを特筆すべき発表として後述

Exhibitionは以下の現地ブースの他、Webのみの展示も行われた。

・現地展示会場171ブース。うち、**日本の企業（アメリカ法人を含む）および研究機関は以下の16機関**（以下はブース番号順）ただし、通常であれば多数の説明員がブースにいるが、今回はほぼ無人の状態のところも多かった。

理化学研究所、高エネルギー加速器研究機構、情報通信研究機構、日東工器、日本電気、プリファードネットワーク、ニッタ、東京大学（生産技術研究所、情報基盤センタは別で計3ブース）、東京工業大学、富士フィルム、海洋研究開発機構、宇宙航空研究開発機構、筑波大学、大阪大学

3. Technical Papers Award

Award Sessionにおいて以下のAwardが発表された。※赤字の発表は特筆すべき発表として後述

区分	タイトル	著者	著者所属
Best Paper	Revealing Power, Energy and Thermal Dynamics of a 200PF Pre-Exascale Supercomputer	Woong Shin	Oak Ridge National Laboratory
Best Student Paper	Efficient Large-Scale Language Model Training on GPU Clusters	Deepak Narayanan	Stanford University
Best Poster	Enabling Combustion Science Simulations for Future Exascale Machine	Jon S. Rood	National Renewable Energy Laboratory
Gordon Bell Prize*	Closing the “Quantum Supremacy” Gap: Achieving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputing	Yong (Alexander) Liu, Xin (Lucy) Liu, Fang (Nancy) Li, Yuling Yang, Jiawei Song, Pengpeng Zhao, Zhen Wang, Dajia Peng, Huarong Chen, Haohuan Fu, Dexun Chen, Wenzhao Wu, Heliang Huang and Chu Guo	Zhejiang Lab in Hangzhou, National Supercomputing Center in Wuxi, Tsinghua University, Shanghai Research Center for Quantum Sciences
Gordon Bell Special Prize in HPC-Based COVID-19 Research**	Digital Transformation of Droplet/Aerosol Infection Risk Assessment Realized on “Fugaku” for the Fight Against COVID-19	Kazuto Ando, Rahul Bale, ChungGang Li, Satoshi Matsuoka, Keiji Onishi, Makoto Tsubokura	RIKEN

* ACMが毎年授与する、高性能コンピュータ分野で顕著な業績のあったシステムや技術を讃える賞。

** 高性能コンピューティングの使用によるCOVID-19のパンデミックの理解に向けた優れた研究成果を評価するために、2020年と2021年に授与される。

4. キーノート、招待講演、受賞講演

■ キーノート

講演タイトル	講演者
Computing and the Humanities	Vint Cerf

■ 招待講演

講演タイトル	講演者
Let the Data Flow!	Kunle Olukotun
Quantum Internet: From a Physics Experiment to a Quantum Network System	Stephanie Wehner
High Performance Convergence Computing	Yutong Lu
HiPEAK: 17 Years of Growing an HPC Community in Europe	Koen De Bosschere
The First 'Exascale' Supercomputer, and Its Application to Society 5.0 to Transform Society for Sustainability and Beyond	Satoshi Matsuoka

■ 受賞講演

受賞区分	講演者/所属
Ken Kennedy Awards	Dr. David Abramson/ University of Queensland
Sidney Fernbach Awards	Prof. David Bader/ New Jersey Institute of Thechnology
Test of Time Award	George Karypis and Vipin Kumar

5. SCおよび特筆すべき発表の概要

● SC領域での世界の流れ

スーパーコンピュータの世界的な傾向として、今回TOP500の上位に変動はなかったが、**来年にはアメリカのエクサフロップスマシンが登場し、新たな時代に突入することが期待**される。また、中国のスーパーコンピュータもすでに複数のシステムがエクサフロップスを超えているとの発表も行われており、目が離せない。また、**応用分野の点でも、従来の科学技術計算に加えて、ディープラーニングや量子計算などの新たな分野でのスーパーコンピュータの利用が活発**となってきており、今後も適用分野の拡大が期待される。

● SCの技術動向

TOP500の発表からもわかる通り、当初2020年頃と予測されたエクサフロップを超えるスーパーコンピュータは、2021年でも実現されていない。2013年頃まではスーパーコンピュータの性能は10年で1000倍向上していたのに対し、それ以降は10年で100倍程度とその速度は鈍化している。これは半導体プロセス技術の進歩の鈍化と密接に関連しているが、一方では着実に性能向上を果たしているとみることもできる。**半導体プロセスの高密度技術は、いくつかの問題点を指摘されつつも、微細化の進歩はまだしばらくは続きそうであり、さらに、3次元実装など実装技術による高密度化との組み合わせにより、これからもスーパーコンピュータの速度向上は続きそうである。**

● 将来的な研究開発の方向性

トランジスタが小さくとも電力密度は一定を保つというデナード則は、微細化による漏れ電流の増大や入力電圧がスケールしないことにより、すでに終焉しており、高密度にするとそれだけ電力密度も増大してしまう。今後とも、スーパーコンピュータの開発の一番の問題は電力問題であり、**いかにして電力効率を向上させていくかが重要**である。また、今回の論文賞を受賞した発表のように、システムの電力効率を向上させるだけではなく、冷却システムやジョブスケジューラと連携した**運用による電力効率の向上がより重要**となるであろう。また、**異なるシステムでいかにアプリケーションのポータビリティを上げていくかが重要な課題**となっていくであろう。

5-1. 特筆すべき発表の概要

主にスーパーコンピュータのシステム開発の観点から以下の発表について注目した。

1. 性能の観点から：TOP500 BOF：本発表は、スーパーコンピュータの性能に関するランキングに関する発表であり、今回はTOP10に大きな変化はなかったが、**中国の動きなど注目すべき発表**があった。
2. 電力効率の観点から：Gree500 BOF：本発表は、スーパーコンピュータの電力効率に関するランキングに関する発表であり、**近年電力効率の向上が目覚ましく、注目**していた。
3. 運用の観点から：Summitの運用データ：本発表は、論文賞を受賞した発表であり、現在世界第二位のスーパーコンピュータであるSummitの電力などの運用データを解析した論文である。**今後の大規模システムで電力効率の高い運用を目指す意味で非常に興味深い。**
4. 最新システムの観点から：OceanLightによるQCRシミュレーション：本発表はGordon Bell賞を受賞した発表であり、中国の最新スーパーコンピュータを用いた大規模シミュレーションで単精度であるが1EFLOPSを超えた性能を報告している。この**OceanLightはTOP500にはエントリーしておらず、詳細が明らかになるのは初めてであり、注目した。**

次スライド以降に注目した発表の詳細を紹介する。

5-1-1. Top500 BOF (1/2)

(1) 本発表が応用される分野、製品

本発表のランキングはLINPACKという大規模密行列連立方程式を解くベンチマークにより世界のスーパーコンピュータの性能によるものであり、演算能力が特に必要となる科学技術計算で使われる。

(2) 当該分野における技術ニーズ

Top500は演算性能を競うランキングであり、微細化および3次元実装による高性能化、および多数のノードを効率的に使う技術ニーズがある。

(3) 本発表の概要

富岳が4期連続でTop1となったほか、Top10は10位を除き前回と変化がなかった。

Rank	Site	Manufacturer	Computer	Country	Nodes	Rmax (MFLOPS)	Power (MW)
1	RIKEN Center for Computational Science	Fujitsu	Fugaku Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Japan	7,630,848	442.0	29.9
2	Oak Ridge National Laboratory	IBM	IBM Power System, P9 22C 3.07GHz, Mellanox EDR, NVIDIA GV100	USA	2,414,592	148.6	10.1
3	Lawrence Livermore National Laboratory	IBM	Sierra IBM Power System, P9 22C 3.1GHz, Mellanox EDR, NVIDIA GV100	USA	1,572,480	94.6	7.4
4	National Supercomputing Center in Wuxi	NRCPC	Sunway TaihuLight NRCPC Sunway SW26010, 260C 1.45GHz	China	10,649,600	93.0	15.4
5	NERSC - Lawrence Berkeley National Laboratory	HPE	Perlmutter HPE Cray EX235n, AMD EPYC 64C 2.45GHz, NVIDIA A100, Slingshot-10	USA	761,856	70.9	2.59
6	NVIDIA Corporation	NVIDIA	Selene DGX A100 SuperPOD, AMD 64C 2.25GHz, NVIDIA A100, Mellanox HDR	USA	555,520	63.5	2.65
7	National University of Defense Technology	NUDT	Tianhe-2A ANUDT TH-IVB-FEP, Xeon 12C 2.2GHz, Matrix-2000	China	4,981,760	61.4	18.
8	Forschungszentrum Jülich (FZJ)	Atos	JUWELS Booster Module BullSequana XH2000, AMD EPYC 24C 2.8GHz, NVIDIA A100, Mell. HDR	Germany	449,280	44.1	1.7
9	Eni S.p.A	Dell EMC	HPC5 PowerEdge C4140, Xeon 24C 2.1GHz, NVIDIA T. V100, Mellanox HDR	Italy	669,760	35.5	2.25
10	Azure East US 2	Microsoft Azure	Voyager-EUS2 ND96amsr_A100_v4, AMD EPYC 48C 2.45GHz, NVIDIA A100, Mell. HDR	USA	253,440	30.1	



(資料) SC21 発表資料

5-1-1. Top500 BOF (2/2)

(4)技術的に特筆すべき点

技術発表ではないため特になし

(5)本発表が注目される理由

今回発表のランキングでは、米国のEFLOPS(Exa FLOPS*¹)計算機がまだイントール中であったこと、中国からの申請がなかったランキングであり、次回には変動があると思われる。

特に注目されるのは、中国は既に2つのEFLOPSを超えるスーパーコンピュータ(Tianhe-3, OceanLight)が開発されており、もう一つももうじき完成するといわれている。

これらの計算機はGordon Bell賞へはエントリーしてFinalistに選ばれているので完成しているのは確かである。

(6)実用化までの課題

既に実用化されたシステムのランキングのため特になし

*1 Exa FLOPS: 浮動小数点演算を一秒間に 10^{18} 回実行可能な性能を示す。
cf. Peta 10^{15} , Tera 10^{12} , Giga 10^9

ATIP's More Realistic China Top100?										
Rank	Developer	Type of System	Site	Installed	Application	#Cores	Rmax (PF)	Rpeak (TF)	Lpk Eff (%)	
1	National U of Defense Tech	Tianhe-3, Feiteng ARM CPU & accelerator	NSCC Tianjin	2021	Supercomputing center	N/A	~1300	1700000	76	
2	Nat Res Cent of Parallel Comp Eng & Tech	Sunway Pro OceanLight, 98304*Sunway SW39010 390C 2.2 GHz, autonomic network	Qingdao Marine	2021	Supercomputing center	4E+07	1050	1333000	79	
3	A server supplier	Mainframe system of a network company, CPU+GPU heterogeneous many-core processor	Network company	2021	Computing power service	285000	125.0	240000.0	52.1	
4	Sugon	Silcon Cube immersive cooling supercomputer CPU+GPU heterogeneous many-core processor	Chengdu Supercomputer Center	2020	Supercomputing center	N/A	~98.0	170000.0	57.6	
5	Nat Res Cent of Parallel Comp Eng & Tech	Sunway TaihuLight, 40960*Sunway SW26010 260C 1.45 GHz, autonomic network	NSCC Wuxi	2016	Supercomputing center	1E+07	93.0	125436.0	74.2	

(資料) SC21 発表資料

5-1-2. Green500 BOF (1/2)

(1) 本論文が応用される分野、製品

Top500と同じくLINPACKを用いたベンチマークであるが、性能ではなく、電力当たりの性能（GFLOPS*2/W）という指標でランキングを行ったもの。演算能力が必要であり、消費電力の制約がある分野で活用される。

(2) 当該分野における技術ニーズ

超高速・低消費電力の演算装置・システムが必要とされている。

(3) 本発表の概要

Preferred NetworksのMN-3が前回に引き続きTOP1となったが、性能は30GF/Wから40GF/Wへと3割ほど向上させている。また、2位には最新のGPUとAMD CPUを組み合わせたシステムが前回の1位を超える性能を達成しているなど、競争と性能向上が著しい。

Green500 Rank	GFLOPS/W	Name	Site	Computer
1	39.38	MN-3	Preferred Networks	MN-Core Server, Xeon Platinum 8260M 24C 2.4GHz, Preferred Networks MN-Core, MN-Core DirectConnect
2	33.98	SSC-21 Scalable Module	Samsung Electronics	Apollo 6500 Gen10 plus, AMD EPYC 7543 32C 2.8GHz, NVIDIA A100 80GB, Infiniband HDR200
3	31.54	Tethys	NVIDIA Corporation	NVIDIA DGX A100 Liquid Cooled Prototype, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 80GB, Infiniband HDR
4	30.80	Wilkes-3	University of Cambridge	PowerEdge XE8545, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 80GB, Infiniband HDR200 dual rail
5	29.52	HiPerGator AI	University of Florida	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR
6	29.05	Snellius Phase 1 GPU	SURF	ThinkSystem SD650-N V2, Xeon Platinum 8360Y 36C 2.4GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR
7	27.37	Perlmutter	DOE/SC/LBNL/NERSC	HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10
8	27.21	Karolina, GPU partition	IT4Innovations National Supercomputing Center, VSB-Technical University of Ostrava	Apollo 6500, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR200
9	26.96	Meluxina - Accelerator Module	LuxProvide	BullSequana XH2000, AMD EPYC 7452 32C 2.35GHz, NVIDIA A100 40GB, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite
10	26.20	NVIDIA DGX SuperPOD	NVIDIA Corporation	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband

The Green500 BoF, SC'21, Nov 2021

*2 GFLOPS: 浮動小数点演算を一秒間に10⁹回実行可能な性能を示す。
cf. Exa 10¹⁸, Peta 10¹⁵, Tera 10¹²

(資料) SC21 発表資料

5-1-2. Green500 BOF (2/2)

(4)技術的に特筆すべき点

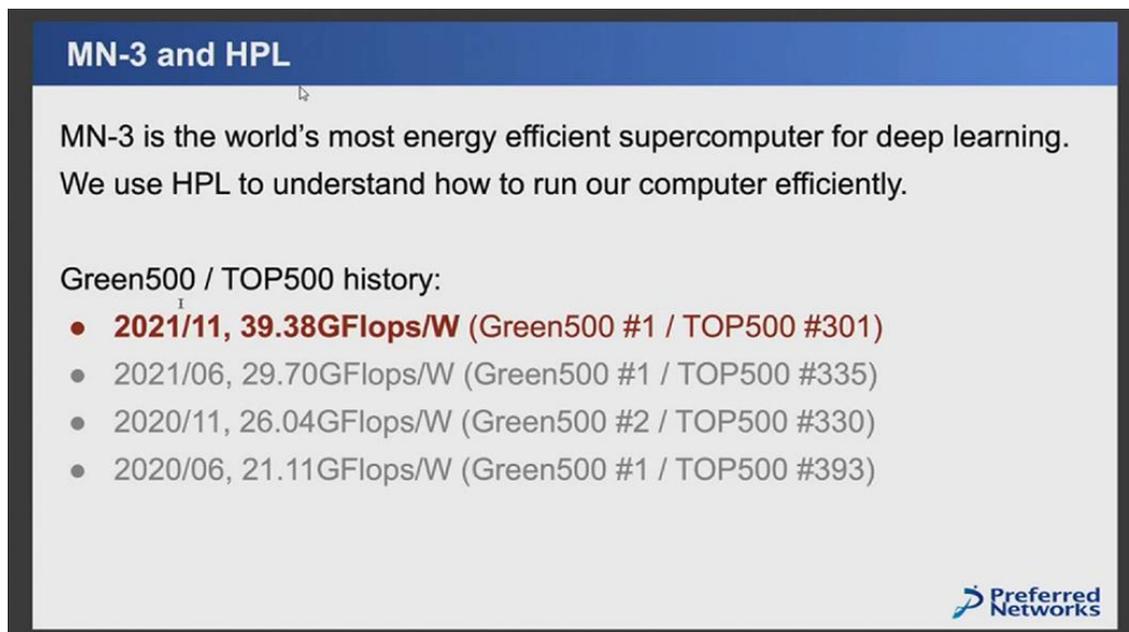
技術発表ではないため特になし

(5)本発表が注目される理由

MN-3は2020/6からGreen500に登録されており、**同一のハードウェアで性能および電力効率を向上させ、最初に比べて約2倍の電力効率を実現している。**発表によると、ソフトウェアの改良、ネットワークの改良、電力制御の改良により実現されている。

(6)実用化までの課題

既に実用化されたシステムのランキングのため特になし



MN-3 and HPL

MN-3 is the world's most energy efficient supercomputer for deep learning.
We use HPL to understand how to run our computer efficiently.

Green500 / TOP500 history:

- **2021/11, 39.38GFlops/W** (Green500 #1 / TOP500 #301)
- 2021/06, 29.70GFlops/W (Green500 #1 / TOP500 #335)
- 2020/11, 26.04GFlops/W (Green500 #2 / TOP500 #330)
- 2020/06, 21.11GFlops/W (Green500 #1 / TOP500 #393)

Preferred Networks

5-1-3. Revealing Power, Energy and Thermal Dynamics of a 200PF Pre-Exascale Supercomputer (1/2)

(1) 本発表が応用される分野、製品

演算能力が特に必要となる科学技術計算や世界的な気候変動シミュレーションなどで使われる。

IBMは、宇宙の起源をさぐるための超新星への理解、癌研究の進展、新材料の開発といった研究に貢献すると謳っている。

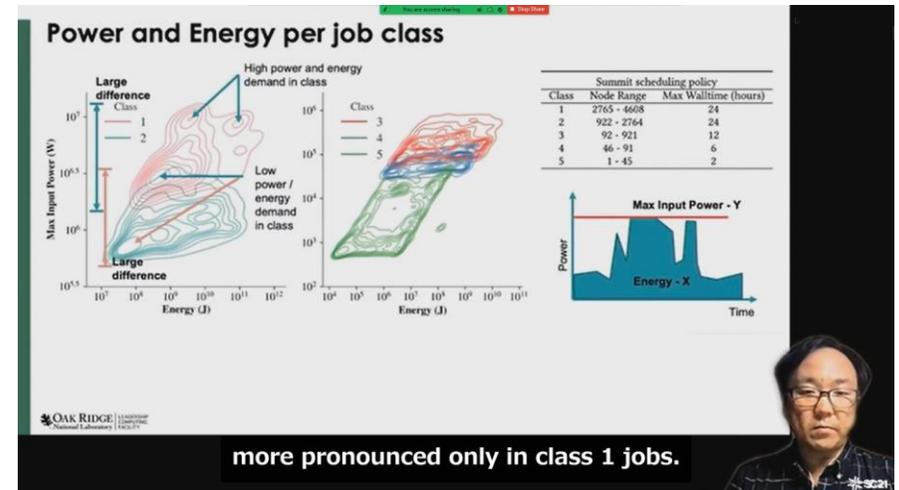
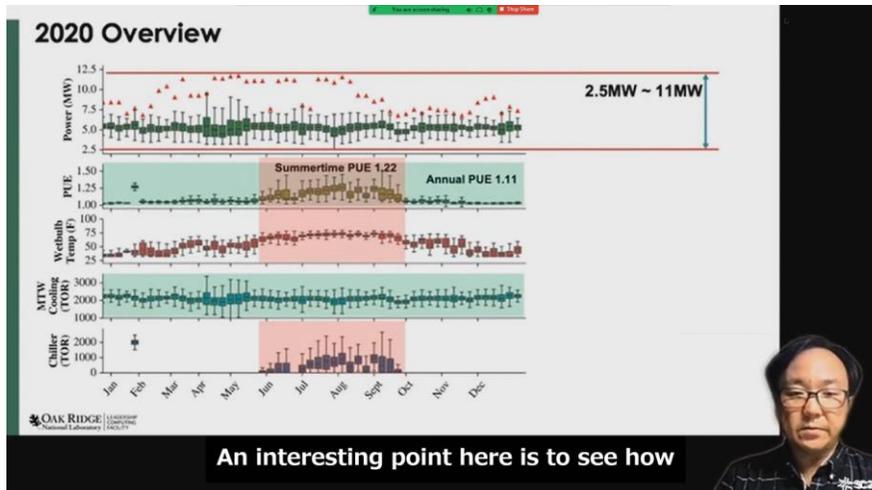
(2) 当該分野における技術ニーズ

多数のCPU、GPUを組み合わせる効率的に稼働させる技術、多数のノード間的高速データ転送技術が必要となる。また、省電力化技術も必要とされる。

(3) 本発表の概要

現在TOP500で2位であり、**ピーク性能200Pflops**である。**4626ノード**からなり、各ノードは**2台のIBM Power9 CPU**と**6台のNVIDIA Tesla V100 GPU**からなる。ピーク電力は**13MW**で、**14.7GFLOPS/W**で2021/7のGreen500では11位であった。

システム全体の電力供給システムや、冷却システムの詳細を紹介するとともに、**電力や冷却水の温度などの実際のモニタリングデータ**を示して、運用時のGF/Wによる電力効率などの時間変化などを示して、**電力の時間変化や、夏場と冬場でのPUE（施設の電力効率を示す指標）の違いなどを詳細に評価している。**また、**ジョブを利用ノード数によりクラス分けをして、各ジョブクラスにおける電力の最小・最大・差分などを評価している。**ジョブの要求するエネルギーに合わせて運用制御することが電力効率向上に有効である。

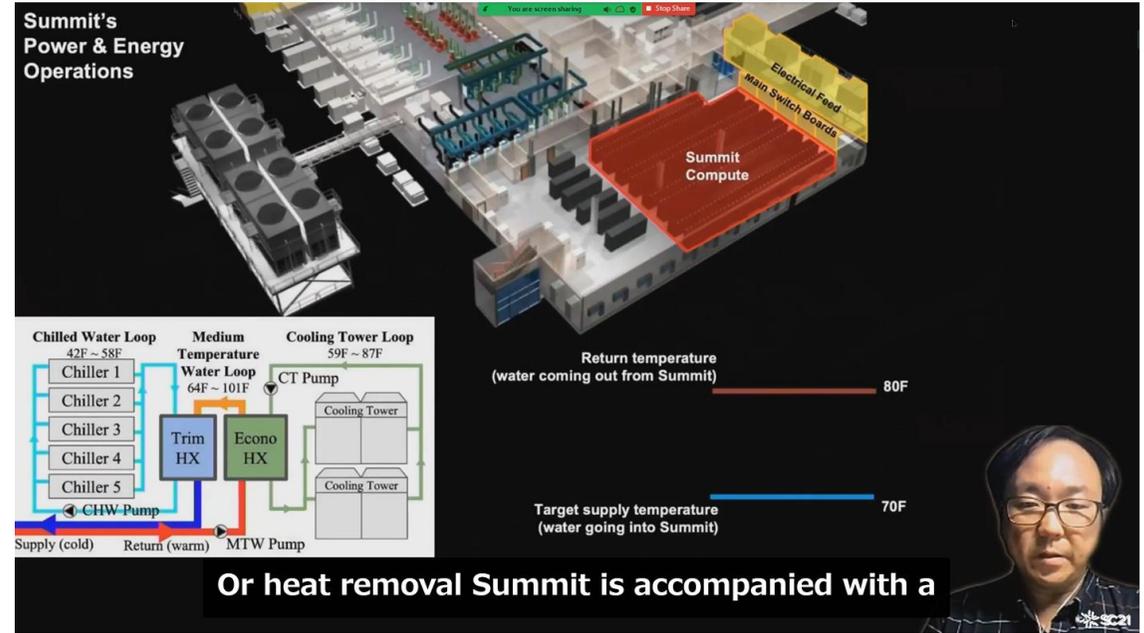


(資料) SC21 発表資料

5-1-3. Revealing Power, Energy and Thermal Dynamics of a 200PF Pre-Exascale Supercomputer (2/2)

(3)本発表の概要(続き)

GPUなどノードは水冷であり、2つの冷却水ループがあり、各ノードの冷却水は入力温度70F(21℃)、出力温度80F(27℃)と富岳などと比べるとやや高い温度設定となっている。通常は蒸発冷却で電力を抑え、夏場(年間の20%)のみ2次冷却を行っている。



(4)技術的に特筆すべき点

技術発表ではないため特になし

(5)本発表が注目される理由

Oak Ridge National Laboratoryに導入されているSummitの電力や性能などに関して詳細に評価を行った論文として注目される。運用時の各種モニタリングデータを長期に保存しておき、今後の運用に役立つ仕組みの構築と実際の考察を緻密に行っている。Best Paperに選ばれた論文である。

(6)実用化までの課題

既に実用化されたシステムのため特になし

(資料) SC21 発表資料

5-1-4. Closing the “Quantum Supremacy” Gap: Achieving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputer (1/4)

(1) 本発表が応用される分野、製品

演算能力が特に必要となる科学技術計算や世界的な気候変動シミュレーションなどで使われる。

(2) 当該分野における技術ニーズ

超高速・低消費電力の演算装置・システムが必要とされている。

(3) 本発表の概要

量子コンピュータの回路を最新Sunway(中国の最新スーパーコンピュータOceanLightと思われる)で高速にシミュレーションを行い**単精度演算で1.2EFLOPS、混合精度演算で4.4EFLOPSを達成**した発表である。Summitスーパーコンピュータを使用してGoogle Sycamore量子プロセッサ(53キュービット)の**ランダム量子回路をシミュレートする最新の取り組みでは実行に10,000年かかると推定されていたが、本発表ではこれを304秒に短縮した。**

OceanLightのプロセッサは、2016年6月のTOP500で一位となったTaihuLightのプロセッサを進化させたもので、コア数を1.5倍、SIMD幅を2倍、クロックを1.5倍でプロセッサ性能を4.5倍、システムのノード数を2.6倍として、全体で12倍のピーク性能1.5PFLOPSを実現しているらしい。基本アーキテクチャは踏襲しているため、ソフトウェアなどは継続して利用できるようである。

5-1-4. Closing the “Quantum Supremacy” Gap: Achieving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputer (2/4)

(3)本発表の概要 (続き)

本発表での量子計算機のシミュレーションを、これまでのシミュレーション結果と比較したのが下記グラフである。横軸がQbit数、縦軸が必要となるメモリ量で、**これまでのシミュレーションに比べて、大規模なQbit数を極めて少ないメモリで実行**するように、OceanLight向けのアルゴリズム的な工夫が行われている。

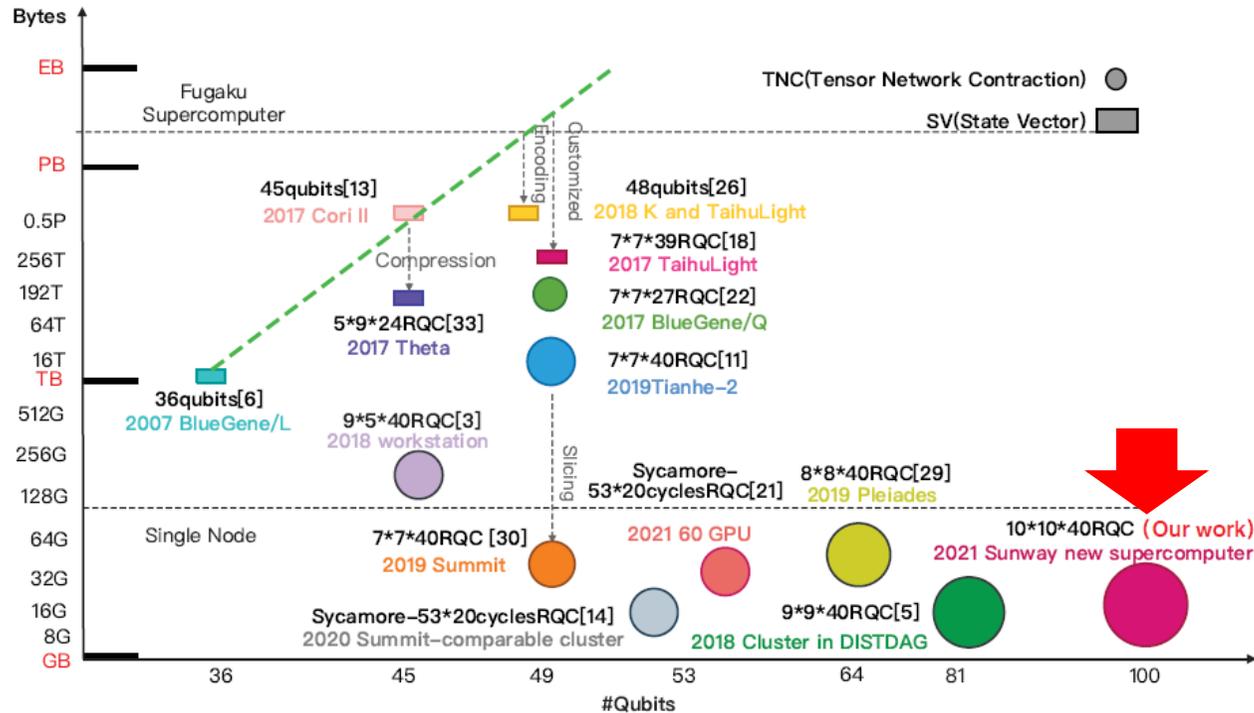


Figure 2: A summary of major classical RQC simulations. The x-axis denotes the number of qubits, while the y-axis shows the corresponding memory space required. The size of the circle/rectangular corresponds to the complexity (depth) of the circuit.

(資料) SC21 発表資料

5-1-4. Closing the “Quantum Supremacy” Gap: Achieving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputer (3/4)

(3)本発表の概要 (続き)

本発表で使われたOceanLightと思われるNew Generation Sunway Supercomputerに関する発表中の下記スライドは1枚だけで簡単な説明のみであった。SW26010Pと名付けられ、TaihuLightのCPUであるSW26010との比較が示された。

Core Groupとして8x8のCoreがメッシュ接続されている構成はSW2010と同じであるが、Core Group数が4から6に増加し、コアのローカルメモリが4倍に、メインメモリも3倍になりメモリバンド幅も2倍強となっている。CPUあたりの演算性能は4.6倍となるとともに、半精度演算をサポートしている。全体で107,520CPU (コア数では約42M) からなるため、ピーク性能では1.5EFLOPSとみられる。

発表後の質問で、ソフトウェアについての質問があり、TaihuLightとの互換性があるとの回答があった。

SW26010	SW26010P
4 x core groups (1 MPE + 8x8 CPEs)	6 x core groups (1 MPE + 8x8 CPEs)
64 KB LDM per CPE	256 KB LDM per CPE
32 GB memory 136 GB/s	96 GB memory 307 GB/s
3 Tflops (single/double)	14 Tflops (single/double) 55 Tflops (half)

The diagram illustrates the processor architecture, showing a central loop network connecting six core groups. Each core group consists of a Memory Processing Element (MPE) and a Core Processing Element (CPE) array. The CPE array is an 8x8 grid of city management, processing elements. The MPE is connected to the CPE array via a DDR memory interface. The loop network is connected to the MPE of each core group. The network interface is also shown.

management, processing elements 1. A 8 by 8 array of city

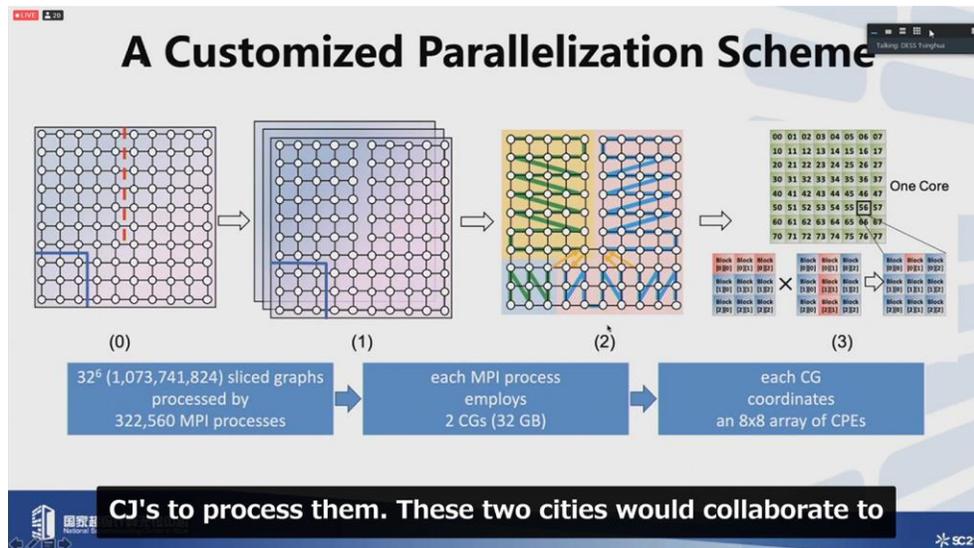
(資料) SC21 発表資料

5-1-4. Closing the “Quantum Supremacy” Gap: Achieving Real-Time Simulation of a Random Quantum Circuit Using a New Sunway Supercomputer (4/4)

(4)技術的に特筆すべき点

本論文で提案している並列化の方式を示しているのが下記スライドである。

まず、元の巨大な2Dテンソルネットワークを、多数 ($32^6 = 1M$) の独立したスライステンソルに分解し、32万個のMPIプロセスに割り当てる。次に、各MPIタスクを2つのCore Groupに割り当てる。さらに、テンソル置換と行列乗算を各Coreに割り当てて実行することにより、大規模並列による高速化を実現している。



(資料) SC21 発表資料

(5)本発表が注目される理由

TOP500に申請していない中国の最新のスーパーコンピュータOceanLight (論文ではnew generation Sunway Supercomputerとだけ示している) を用いた結果ということで注目される。Gordon Bell Prizeに選ばれた論文である。

(6)実用化までの課題

実行時間は他のシステムでは数年かかる計算を304秒で実行できた、ということで、あまり長時間の実行ではなかったようである。TOP500に申請する際に、LINPACKで性能を上げるためには問題サイズを大きくする必要があり、その実行に1時間程度はかかると思われるので、より長時間安定して実行できる必要がある。

2-1-7. RISC-V Summit 2021

目次

1. エグゼクティブサマリ

2. 2021 RISC-V Summit概要

3. プログラム

3-1. 基調講演 (16+showcase 3)

3-2. RISC-V 命令アーキテクチャの拡張 (28)

3-3. 業界応用分野ごとソリューション (18)

3-4. ソフトスタックと開発キット (10)

※括弧内の数字はセッション数を表す

4. 注目発表

4-1. 基調講演

4-2. 命令アーキテクチャ拡張

4-2-1. AI-RISC – Custom Extensions to RISC-V for Energy-efficient AI Inference at the Edge of IoT

4-2-2. Vitruvius

4-2-3. Advanced Interrupt Architecture(AIA) and Advanced CLINT

4-3. セキュリティ

4-3-1. SYSTEMATICALLY SECURING THE RISC-V

4-3-2. The Future of RISC-V Heterogeneous Embedded Virtualization Architectures

4-4. Esperanto社のET-SoC-1

4-5. SiFive

4-6. Andes Technology社

4-7. Microchip PolarFire® SoC FPGA

4-8. EdgeQ

4-9. Mythic & Cudasip

1. エグゼクティブサマリ (1/2)

1-1. 概要および全体サマリ

- RISC-V Summit 2021は2021年12月6日～8日開催。米サンフランシスコでの実会場とバーチャルの**ハイブリッド開催**となった。
- 今回のサミットではRISC-Vが成長のステップに入ったことが浮き彫りになった。RISC-V Internationalは現在**2,478人のメンバー**で構成されており、**ワーキンググループと委員会は約12,000人の個人メンバー**があり、**70カ国・地域以上**が参加している。
- RISC-V Internationalは2020年8月3日、旧組織であるRISC-V Foundationからの移行に伴い、スイスに拠点を置いた。本拠地をスイスに移したのは、スイスが中立的な場所であり、それだけで人々に安心感を与えられるため、と説明している。米国に加え、ヨーロッパ、中国の取り組みが活発。地政学的要因もあると思われる。
- 今回の発表を地域ごとに見ると、
 米国：27、英：5、中国：9、台湾：6、ヨーロッパ：22、日本：2、
 インド：1、イスラエル：1、ロシア：1、アブダビ：1
 となっている。
 - ヨーロッパ：欧州プロセッサ・イニシアチブ（EPI）は、**欧州の半導体独立への道としてRISC-Vに期待**している。
 - 中国：中心はAlibaba, StarFiveだった。DatanLoard, Huawei (UK), CAS(中国科学院)も発表している。国の後押しもあるが、**民間が国家間の覇権争いの影響を危惧してRISC-Vに積極参加**している面もある。
- オープンソースプロセッサであるRISC-Vがかつてないほど急速に採用されている。**2021年、20億のRISC-Vコアが市場に存在**すると見積もっている。
- 応用分野は、エッジコンピューティング、コンシューマIoTデバイス、自動車といった分野に加えAI/機械学習、通信、データセンタ・クラウド、HPCと言った分野でも採用が増えている。**4-1.基調講演で概要を報告、個別に重要なものをピックアップし、4-4.～4-9.で解説**した。
- 標準のISA^{*1}は仕様化されているが、命令セット拡張の議論は活発であり、4-2.で解説。
- **セキュリティ**は大きなトピックであり、4-3.で解説。**レベルごとのアイソレーション、仮想化、暗号化エンジン**がトレンド。

*1 ISA(Instruction Set Architecture): マイクロプロセッサを動作させるための命令語の体系

1. エグゼクティブサマリ (2/2)

命令アーキテクチャ拡張 (4-2のサマリ)

- RISC-Vでは、オプションとして様々な拡張ISAが定義され、更に**ユーザの独自ISA拡張を許容している**。
- 1命令で複数データを扱う方式として、1レジスタを複数データで使用するSIMD方式*2が一般的であるが、RISC-Vでは、レジスタ幅と独立にデータの幅と数を柔軟に指定できるベクトル拡張を定義した。
- さらに、Linux OS実装を想定して定義された割込方式を、リアルタイムOS対応が可能なように拡張している。
- 注目発表では、Virginia大学のAI向け独自ISA拡張、BSC*3のベクトル拡張実装、Western Digital社の割込方式拡張を取り上げた。

セキュリティ (4-3のサマリ)

- セキュリティ強化は必須の流れであり、RISC-V向けにも様々な取り組みがなされている。
- セキュリティ対策は守る対象の価値に見合うコストで実現する必要があり、実用化にはコスト低減が必須である。
- 複雑なシステムでは、脆弱性を完全に排除することは不可能であり、極小化された高信頼なSW/HWで、互いに干渉しないように分割された各領域を守ることで、被害を最小限に抑え込む。
- 注目発表では、NVIDIA社のセキュリティの強化・対策コスト低減、Minho大学のPMP拡張*4、Partitioningとvirtualizationの実装を取り上げた。

*2 前者が本来のSingle Instruction Multiple Data (SIMD) であるが、後者の狭義のSIMDを単にSIMDと呼ぶことが多い

*3 BSC: Barcelona Supercomputing Center

*4 PMP(Physical Memory Protection):物理メモリプロテクション。ソフトウェアによるメモリ領域の保護機構。全てのメモリアクセス (Read/Write/命令フェッチ) に適用される。

2. 2021 RISC-V Summit概要

1. 開催日時

2021年12月6日 - 8日 (3日間)

2. 開催形式

サンフランシスコの実会場とバーチャルのハイブリッド開催

3. 構成

全75セッション

セッション内訳

セッション分野	セッション数
キーノート	19
RISC-V 命令アーキテクチャの拡張	28
業界応用分野ごとソリューション	18
ソフトウェアと開発キット	10



URL <https://riscv.org/event/2021-risc-v-summit/>

3. プログラム

3-1. 基調講演

※赤字の発表は注目発表として後述

講演タイトル	講演者/所属
Welcome & Opening Remarks	Calista Redmond, CEO, RISC-V International
Are the RISC-V Design Freedoms Leading to RISK in Verification Quality?	Larry Lapides, Vice President Sales, Imperas Software Ltd.
Bringing RISC-V to Life: Building our Software Ecosystem	Philipp Tomsich, Founder and Chief Technologist, VRULL GmbH
The Future of RISC-V has No Limits	Dr. Yunsup Lee, Co-Founder & Chief Technology Officer, SiFive
Building Customized Solutions from Open-sources	Xiaoning Qi, Vice President, Alibaba Group
Diversity, Equity, and Inclusion in Open Hardware	Dr. Marjan Radi, Research Technologist Engineer, NVM Systems Architecture, Western Digital & Kim McMahon, Director of Visibility & Community Engagement, RISC-V International
State of the Union	Krste Asanović, Professor, EECS Dept, U.C Berkeley / Chief Architect and Co-Founder, SiFive Inc.
Beefing Up the Datacenter Accelerators	Charlie Su, President and CTO, Andes Technology
Keynote Panel: RISC-V Momentum at Data Center Scale	Balaji Baktha, Ventana (Moderator); Sumit Gupta, Google; Jing Yang, Alibaba; Roger Espasa, Semidynamics Technology Services; Bapi Vinnakota, Open Compute Project ODSA Project Lead
Microchip and the Expanding RISC-V Universe	Ted Speers, Technical Fellow, Microchip
Profiles and Platforms: RISC-V Convergence	Greg Favor, CTO, Ventana Micro Systems
Road Ahead	Mark Himmelstein, CTO, RISC-V International
Is Hardware/software Co-design for Applications Now a Reality with RISC-V?	Kevin McDermott, Vice President Marketing, Imperas Software Ltd
Awards Presentation	Kim McMahon, Director of Visibility & Community Engagement, RISC-V International & Mark Himmelstein, CTO, RISC-V International
Scaling is Failing	Dr. Ron Black, CEO, Codasip
Where is RISC-V Going?	Calista Redmond, CEO, RISC-V International

3-2. RISC-V 命令アーキテクチャの拡張 (1/2)

※赤字の発表は注目発表として後述

講演タイトル	講演者/所属
A Posit Arithmetic Unit Enabled RISC-V Processor	Aneesh Raveendran & Vivian Desalphine, Centre for Development of Advanced Computing, Bangalore, India
Implementing Functionally-safe RISC-V IP for Automotive and Safety Critical Applications	Shubu Mukherjee, SiFive
Open Hardware for the Open Cloud	Daniel Mangum, Upbound
Implementation of an Out-of-order RISC-V Vector Unit	Roger Espasa, SemiDynamics Technology Services
Vitruvius: An Area-Efficient RISC-V Decoupled Vector Accelerator for High Performance Computing	Francesco Minervini & Oscar Palomar Perez, Barcelona Supercomputing Center - BSC
YoC - an Open Operation System for IoT	Vincent Cui, Alibaba
Algorithm Acceleration for RISC-V Processors using High-Level Synthesis	Russell Klein, Siemens EDA
Advanced Interrupt Architecture and Advanced CLINT	Anup Patel, Western Digital & John Hauser, Independent Researcher
Support for Non-Coherent I/O Devices in RISC-V	Greg Favor & David Kruckemyer, Ventana Micro Systems
Efficient Issue Scheduling for Hardware Multithreaded RISC-V Pipeline	Dr. Shlomo Greenberg, Ben Gurion University of the Negev & Sami Shamoon College Engineering, Beer-Sheva
Israel RISC-V Compatible Processor IP by Syntacore: Compact Open-source MCU to Multicore Linux	John Hartley, Syntacore
Lightning Talk: RISQV-HT: A RISC-V Microcontroller Delivering Post-Quantum Secure Encryption - and Hardware Trojans	Alexander Hepp, Technical University of Munich
Accelerating AI and non-AI Workloads with 1000+ Energy-Efficient RISC-V Cores on a Single Chip	Art Swift, Esperanto Technologies
RISC-V Debug in the OS-A Platform	Paul Donahue, Ventana Micro Systems

3-2. RISC-V 命令アーキテクチャの拡張 (2/2)

※赤字の発表は注目発表として後述

講演タイトル	講演者/所属
AI-RISC - Custom Extensions to RISC-V for Energy-efficient AI Inference at the Edge of IoT	Vaibhav Verma, University of Virginia
Exploring the Zce Code-size Reduction ISA Extension	Tariq Kurd, Huawei UK
Webassembly as Managed Runtime VM in Embedded Systems	Stefan Wallentowitz, Munich University of Applied Sciences
Lightning Talk: Functional Gap between RISC-V V and SPIR-V: a Study Case on the Graphics Domain	Abel Bernabeu, Esperanto Technologies
Lightning Talk: A Zero Trust Security Architecture For RISC-V SoC/ Platform	Suresh Sugumar, Technology Innovation Institute
Esperanto's Custom RISC-V ISA Extensions for Energy-Efficient Machine Learning Applications	Jayesh Iyer, Esperanto Technologies
Radiation Hardening and Fault-Tolerance Features of the NOEL-V RISC-V Processor	Jan Andersson, CAES Space System Division, Gaisler
Lightning Talk: Adding H Support to the NOEL-V Microprocessor	Stefano Ribes, De-RISC Project
Lightning Talk: A Secure RISC-V Based SoC for Autonomous UAVs Navigation	Davide Rossi, University of Bologna & Daniele Palossi, IDSIA USI-SUPSI Lugano
Unveiling the SweRV Core EH3	Zvonimir Bandic, Western Digital
Profiles and Platforms	Philipp Tomsich, VRULL & Mark Himmelstein, RISC-V International
Systematically Securing the RISC-V - Secure Foundation for Embedded Functionality	Marko Mitic, NVIDIA
IOPMP Updates: The Protection of IOPMP	Chyun Ku, Andes Technology
Architecture Design for Security: Do's and Don'ts	Gregory T. Sullivan, Dover Microsystems, Inc.

3-3. 業界応用分野ごとソリューション

※赤字の発表は注目発表として後述

講演タイトル	講演者/所属
XiangShan: an Open-source High-performance RISC-V Processor	Yungang Bao, Institute of Computing Technology, Chinese Academy of Sciences (ICT, CAS)
Lightning Talk: Improving Performance of National Crypto Algorithms with Custom Instructions	Alexander Kozlov, CloudBEAR
Lightning Talk: A System Level Verification and Validation Environment using SweRV	Anupam Bakshi, Agnisys, Inc.
Lightning Talk: How to Extend RISC-V to Accelerate AI/ML	Veronia Iskandar, TU Dresden & Dr. William Jones, Embecosm
Lightning Talk: Open-Source RISC-V Cores with Industrial Strength Verification	Simon Davidmann & Lee Moore, Imperas Software
Lightning Talk: De-RISC, the Horizon 2020 Project that will Create the First RISC-V	Paco Gómez-Molinero, fentISS
Lightning Talk: Adding 32-bit Linux Support to ARIANE/CVA6 Open-source Application Core	Sébastien Jacq & Jérôme Quévremont, Thales
Extending RISC-V Instructions for 5G Intelligent RAN Base Stations	Roman Shaposhnik & Kathy Giori, ZEVEDA Inc.
RISC-V on Edge: Porting EVE and Alpine Linux to RISC-V	Suresh Sugumar, Technology Innovation Institute
Lightning Talk: Accelerating Real-World AI Software using the RISC-V Vector Extension	Colin Davidson & Alastair Murray, Codeplay Software
Lightning Talk: Using and Extending RISC-V in an Analog Matrix Processor for Neural Networks	David Luo, Mythic & Dr Zdeněk Přikryl, Cudasip
Hard Real-Time vs High Performance Real-Time Applications on PolarFire SoC	Hugh Breslin, Microchip Technology
Continuous Innovation in Embedded RISC-V Processors	Drew Barbier, SiFive
A Requirements-based Test Suite for the C Standard Library: SuperGuard	Marcel Beemster, Solid Sands B.V.
TEEP (Trusted Execution Environment Provisioning) and Software Updates for Internet of Things (SUIT) on RISC-V	Akira Tsukamoto, AIST
Lightning Talk: Enabling RISC-V Software Ecosystem with VisionFive - an Affordable and Open Source RISC-V Single Board Development Platform with Support for Fedora	Chin Hu Ong, StarFive Technology
Lightning Talk: First Volume Production RISC-V Silicon/SOC to Provide Complete Personal Computing Platform Solution Targeting Mid-range and Edge Computing	Johnson Sun, StarFive
Storage Area Network Acceleration using RDMA / RoCE and RISC-V -	Pu Wang, DatenLord

3-4. ソフトスタックと開発キット

※赤字の発表は注目発表として後述

講演タイトル	講演者/所属
Performance Monitoring in RISC-V using perf	Atish Patra, Western Digital
ACPI for RISC-V: Enabling Server Class Platforms	Sunil V L, Ventana Micro Systems
The Future of RISC-V Heterogeneous Embedded Virtualization Architectures	Sandro Pinto & José Martins, Universidade do Minho
Lightning Talk: Enabling Software Emulation for RISC-V Heterogeneous Cores Architecture	Cui Jin & Ley Foon Tan, StarFive Technology
Sail Specification for RISC-V P-Extension	Bow-Yaw Wang, Academia Sinica, Taiwan & Jenq-Kuen Lee, National Tsing Hua University, Taiwan
BoF: How RISC-V CPU Design Impacts Performance of Copy Function and Network Speed	Akira Tsukamoto, AIST
RISC-V Enterprise Software Ecosystem Readiness	Kumar Sankaran, Ventana Micro Systems
Quantitative Methods for Continuously Improving RISC-V Compilers	Philipp Tomsich, VRULL
Lightning Talk: Performance of TVM AutoScheduler for Andes Vector Processor	I-Wei Wu, Andes Technology
Lightning Talk: Bring Multicore RISC-V and Zephyr RTOS Together	Chun-Wei Shu, Andes Technology

4. 注目発表

4-1. 基調講演：構成メンバーおよび市場（1/2）

・ RISC-Vはどこへ向かっているのか？ - Calista Redmond CEO RISC-V International

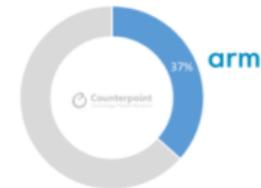
- 今回のサミットではRISC-Vが成長のステップに入ったと報告された。RISC-V Internationalは現在2,478人のメンバーで構成されており、ワーキンググループと委員会は約12,000人の個人メンバーがあり、70か国・地域以上が参加している。
- オープンソースプロセッサであるRISC-Vがかつてないほど急速に採用されている。2021年、20億のRISC-Vコアが市場に存在すると見積もっている。
- RISC-Vは、エッジコンピューティング、コンシューマIoTデバイス、自動車といった分野に加えAI/機械学習、通信、データセンター・クラウド、HPCと言った分野でも採用が増えている（次スライド参照）
- 2025年に半導体IPのトータル市場規模は86億ドルに成長し、
 - IoT向け半導体IPの1/3はRISC-Vになる
 - 自動車向け半導体IPの10%はRISC-Vになると予想されている ※

※Counterpoint Research
Sep 13, 2021
<https://www.counterpointresearch.com/riscv-semiconductor-ip-market-2025/>

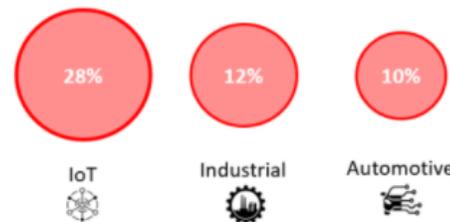
Semiconductor IP market size, 2020 vs 2025



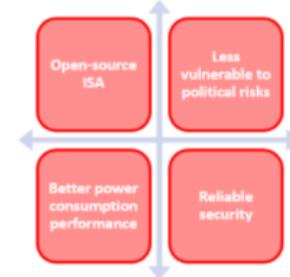
ARM dominates global pure play IP market with 37% share



RISC-V Penetration Rate by 2025



Advantages RISC-V offers



4-1. 基調講演：構成メンバーおよび市場（2/2）

- RISC-Vの採用分野、企業：エッジコンピューティング、コンシューマIoTデバイス、自動車といった分野に加えAI/機械学習、通信、データセンタ・クラウド、HPCと言った分野でも採用が増えている

Edge Computing

A distributed, open architecture decentralizes processing power, reduces latency, and supports IoT performance in low bandwidth environments at low power.

- **Seeed Studio's** new development, the Sipeed MAIX, a RISC-V 64 AI board for Edge Computing makes it possible to embed AI to any IoT device.
- **Micro Magic** announced an incredibly fast 64-bit RISC-V core achieving 5GHz and 13,000 CoreMarks at 1.1V.
- **Western Digital** SweRV Core enables spectrum of compute at the edge

Consumer and IoT devices

- **Huawei** Hi3861 RISC-V board for Harmony OS developers for the IoT market
- **Zepp Health** / Huami wearable manufacturer OS supporting RISC-V Reference Models for RISC-V P extension
- **GreenWaves** ultra-low power GAP9 hearables platform for scene-aware and neural network-based noise reduction.
- **Microchip** released the first SoC FPGA development kit based on the RISC-V ISA.
- **RIOS Lab** announced PicoRio, an affordable RISC-V open source small-board computer.
- **SiFive** world's fastest development board for RISC-V Personal Computers.

Automotive

- **Imagination Technologies** GPU can be linked together by a RISC-V core for ASIL-B level designs with ISO26262 safety critical certification.
- **IAR Systems** extended the functional safety version of its Embedded Workbench software tool chain to the RISC-V core of Nsitex, subsidiary of automotive leader Denso.
- **Renesas and SiFive** jointly develop next generation RISC-V SoCs for automotive
- **Renesas and NSI-TEXE** announce automotive SoC with RISC-V-based parallel co-processor
- **Europe's GaNext project** to simplify designing power converters with GaN power semiconductors while improving efficiency and compactness for systems such as electric vehicle chargers.

AI / ML

Artificial intelligence spans many areas from Industrial IoT to financial

- **Esperanto** Emerges From Stealth With 1,000-Core RISC-V AI Accelerator.
- **StarFive** released the world's first RISC-V AI visual processing platform
- **Andes** released superscalar multicore and L2 cache controller processors.
- **NVIDIA CUDA** support on the Vortex RISC-V GPGPU enables scaling from 1-core to 32-core GPU based on RV32IMF ISA with OpenCL 1.2 graphics API support.

Telecom & Communications

- **Andes** 64-bit RISC-V processor has been adopted by SK Telecom for the development of AI products.
- **Alibaba** PLCT Lab has ported Android 10 onto its in-house 64-bit RISC-V core emulated in QEMU
- **Google** Pixel 6 Titan M2 in-house designed RISC-V processor, with extra speed and memory, and even more resilient to advanced attacks.

Data Center Could HPC

- **Alibaba** RISC-V Xuantie processor line including four open sourced processors for cloud and edge servers
- **Imagination** RISC-V CPU family, for both the discrete CPU and heterogeneous computing markets
- **Seagate** hard disk drive controller with high-performance RISC-V CPU.
- **Ventana** performance chiplet approach to data center SoC design
- **Intel** Nios processor based on RISC-V, designed for performance.

High Performance Computing

- **European Processor Initiative** RISC-V accelerator with first chip Sep 2021
- **Technical University of Munich (TUM)** quantum cryptography chip for quantum computing security demands
- **Tactical Computing Labs** HPC-centric software test suite for GCC and LLVM
- **Cortus** is developing a high-performance RISC-V Out-of-Order processor core for the European eProcessor project.
- **De-RISC** market-ready HW-SW platform for a multi-core RISC-V system-on-chip for safety critical aerospace applications

出典：図は2021 RISC-V Summitの発表資料を引用

4-1. 基調講演：RISC-V仕様および互換性

・Road Ahead - Mark Himmelstein, CTO, RISC-V International

- RISC-Vの仕様は、ISA仕様、デバッグ仕様、およびトレース仕様の3つと、互換性テストフレームワークがある。

<https://riscv.org/technical/specifications/>

- 2021年、RISC-Vメンバーは、15の新しい仕様と40を超える拡張機能を承認した。新しい仕様は以下の6分野に重点をおいている。

No	分野
1	オーディオ、ビジュアル、音声のML推論
2	複雑なデータ配列処理
3	マルチタイプのハイパーバイザーと仮想マシン（VM）の実装
4	自動車、データセンター、および産業用制御に関連するクラウドネイティブアプリケーション
5	IoTと組み込みデバイス
6	暗号化ハッシュとブロック暗号化によるセキュリティ

4-2. 命令アーキテクチャ拡張

4-2-1. AI-RISC - Custom Extensions to RISC-V for Energy-efficient AI Inference at the Edge of IoT --- Virginia大学 (1/3)

(1) 本発表が応用される分野、製品

IoTのEdge側でのAI Inferenceを必要とする分野・製品

(2) 当該分野における技術ニーズ

- ・ネットワークを介さない処理で、ネットワークの遅延・障害の回避、プライベートデータ保護・秘匿の実現
- ・Edge側の限られたリソースでのリアルタイム処理

(3) 本発表の概要

- ・ IoTのEdge側でのAI Inferenceの必要性
- ・ AI応用のための命令拡張、AIタスクを細粒度でオフロードできるアクセラレータ
- ・ HW/SWコデザイン (TVM、Synopsys ASIP Designer、LLVMなどを活用)
- ・ 命令拡張によって+27%の面積で、GEMV*⁵ kernelで5.68倍、MLperf Tiny*⁶で4.41倍の性能を達成

(4) 技術的に特筆すべき点

AI処理を加速するための命令定義と、それを活用するためのコンパイラ、ライブラリの整備

(5) 本発表が注目される理由

独自命令拡張に伴うコンパイラの問題の解決策を考えている。

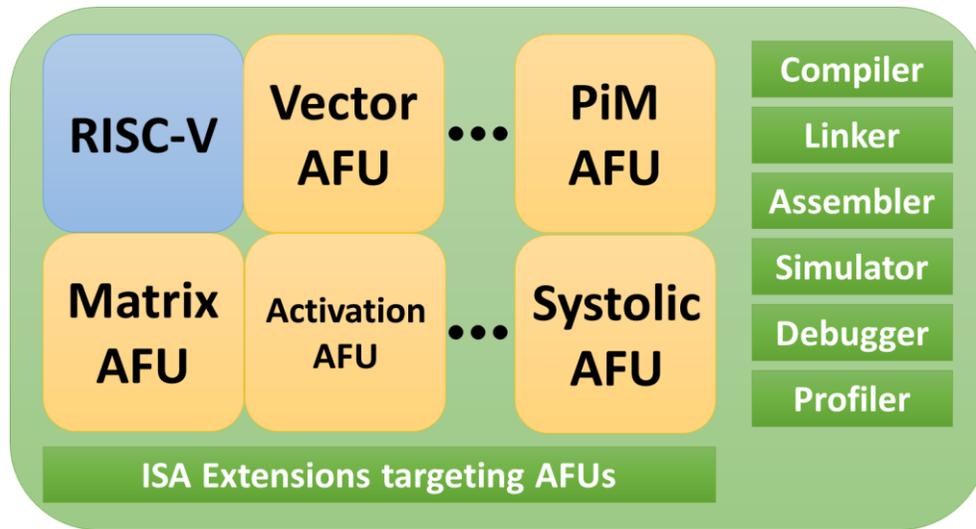
(6) 実用化までの課題

研究レベルの発表であり、企業との連携など実用化の枠組みが必要

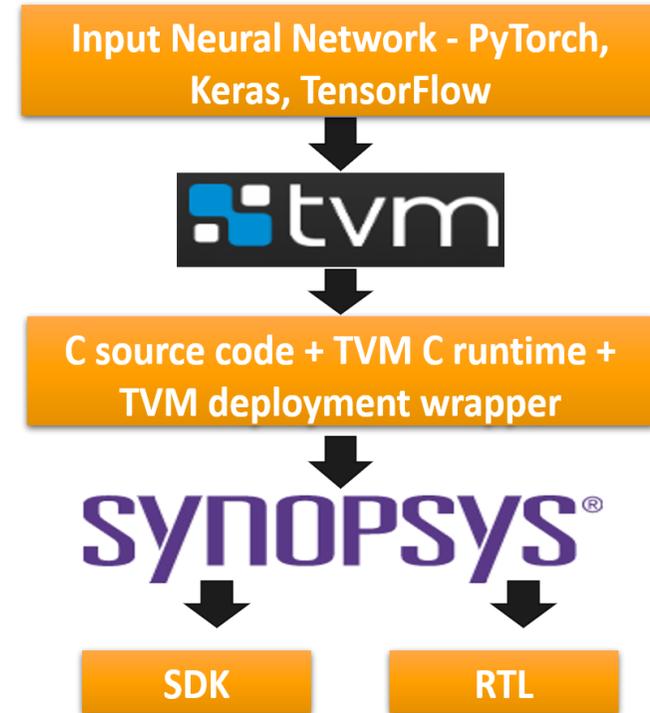
*5 GEMV(General Matrix-Vector Multiply): 行列とベクトルの積 GEMV kernelは行列とベクトルの積のベンチマーク

*6 組込デバイス向け深層学習ベンチマーク

4-2-1. AI-RISC - Custom Extensions to RISC-V for Energy-efficient AI Inference at the Edge of IoT --- Virginia大学 (2/3)

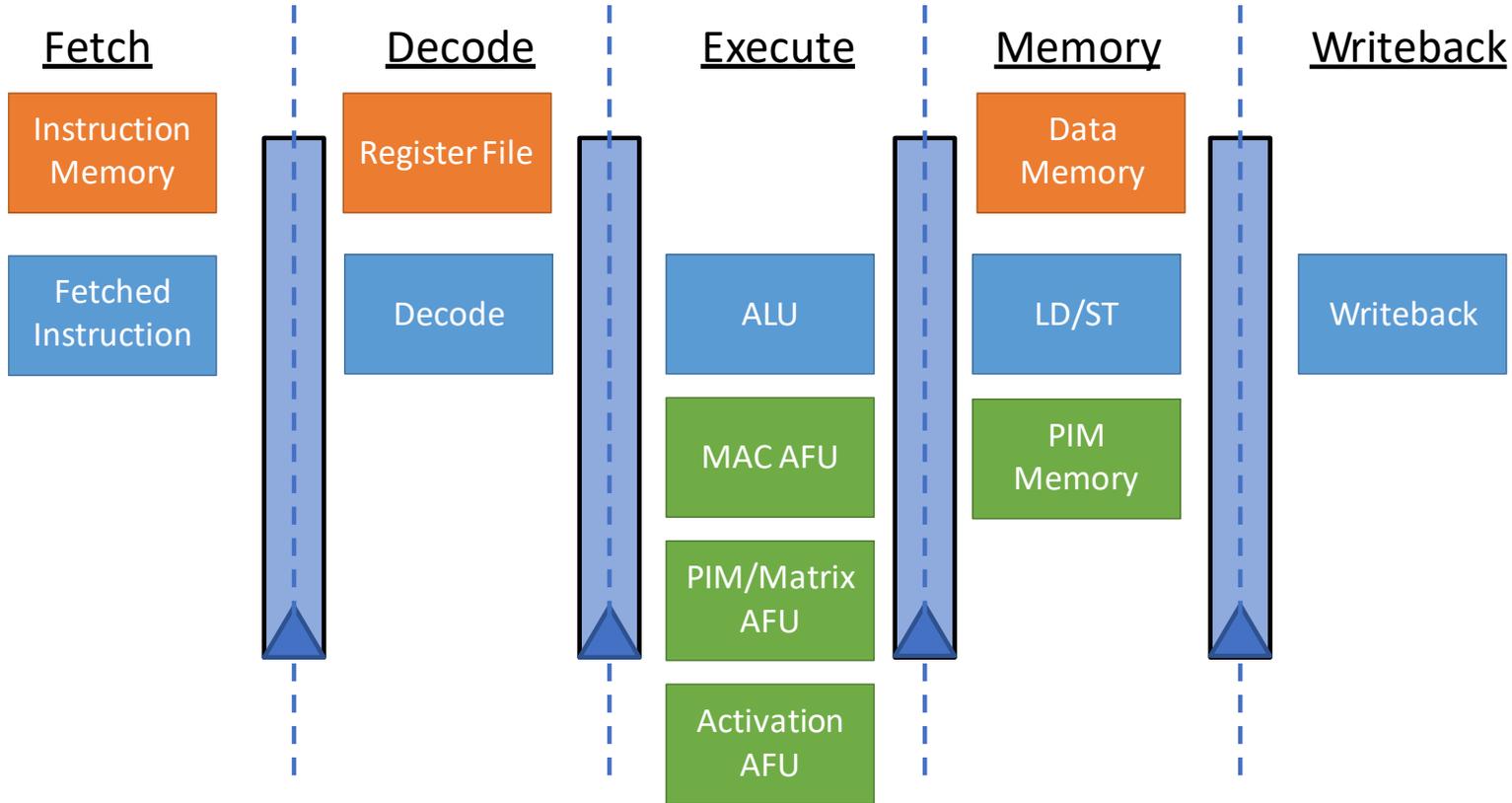


- AI向け拡張命令を処理するHW/SW
 - 様々なAI Functional Unit (AFU) を実装
 - AFU向けの拡張命令
 - 拡張命令対応のツール群



- HW/SWコードデザインフロー

4-2-1. AI-RISC - Custom Extensions to RISC-V for Energy-efficient AI Inference at the Edge of IoT --- Virginia大学 (3/3)



- パイプラインステージに組み込まれた密結合AFU

出典：図は2021 RISC-V Summitの発表資料を引用

4-2-2. Vitruvius --- BSC(Barcelona Supercomputing Center) (1/3)

(1) 本発表が応用される分野、製品

スーパーコンピュータ (Exascaleがターゲット)

(2) 当該分野における技術ニーズ

大規模高並列倍精度処理の高性能化

(3) 本発表の概要

- RISC-VのVector Extension (version 0.7.1) を実装
- **European Processor Initiative (EPI)** の中核部品として開発
- GF 22FDX® 22nm FD-SOIを使い2021年6月にテープアウト
- **最大256要素の倍精度vector処理、最大16 DP-FLOPS/cycle**
- Reduction処理の高速化
- 構成可変なデザイン (可変長vector、可変pipeline段数)

(4) 技術的に特筆すべき点

In-order演算*8とout-of-orderメモリアクセス*9の組合せ

(5) 本発表が注目される理由

- **RISC-Vの最初のVector Extension実装例**
- 別組織の発表でVector演算のout-of-order処理に挑戦して、その大変さを説明しているが、逆にそれを回避することで高効率実装を実現している。

(6) 実用化までの課題

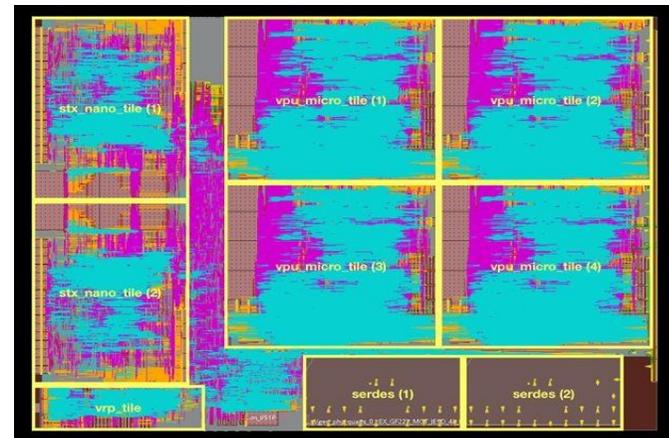
先端プロセスでの実装、システム構築

*7 EPAC(European Processor Accelerators): EPIが作成した電力効率の高い、高スループットのアクセラレータコア

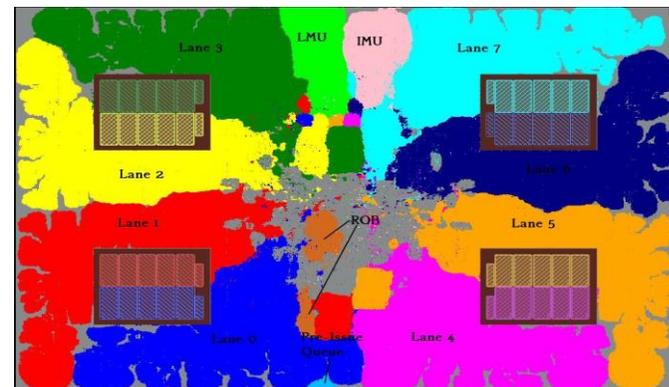
*8 プログラムに記述された順に行う演算

*9 プログラムに記述された順番から外れて行うメモリアクセス

出典：図は2021 RISC-V Summitの発表資料を引用



EPAC*7 layout highlighting the accelerators with 25 mm²



Physical Design

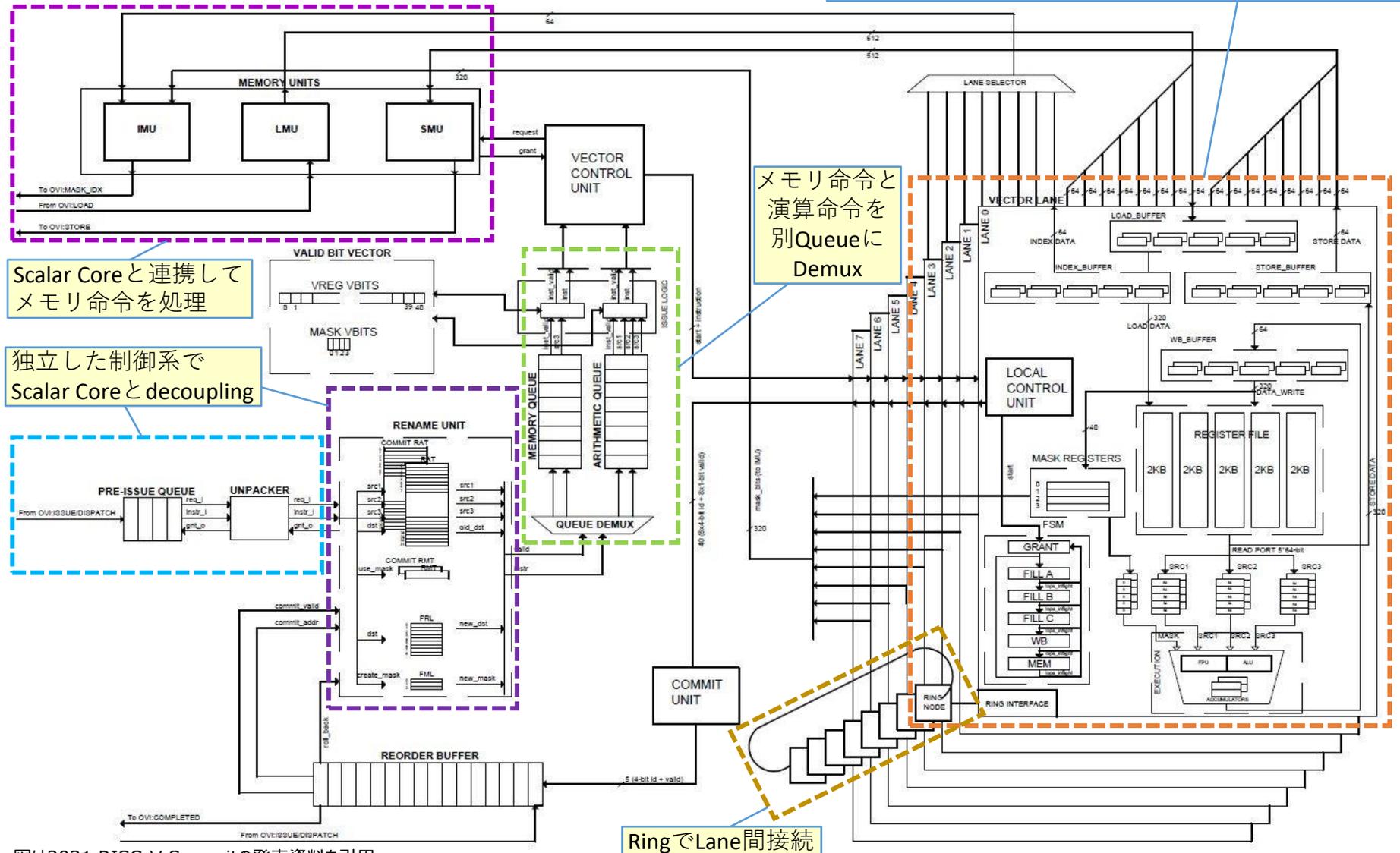


456

4-2-2. Vitruvius --- BSC(Barcelona Supercomputing Center) (2/3)

演算命令はin-order処理であり、Lane毎の制御による処理のoverlapとLane内の単純なSIMD処理によって、効率的に高速化

Vector Acceleratorブロック図



出典：図は2021 RISC-V Summitの発表資料を引用

4-2-2. Vitruvius --- BSC(Barcelona Supercomputing Center) (3/3)

■ 性能評価

- Vector化性能 (ピーク性能は16 FLOPS/cycle, 20 GFLOPS)

Benchmark	DP-FLOP/cycle *10	DP-GFLOPS(@1,25GHz)
Matmul 256x256	15.1	18.9
Jacobi-2D	7.5	9.4
Black-Scholes	5.4	6.75
Daxpy	4.1	5.125
FFTW	2.6	3.25

- 性能比較

Vector Unit	Technology	Frequency [GHz]	Area [mm ²]	Peak Performance [DP-FLOP/cycle]	Efficiency [DP-FLOPS/mm ²]
SX-Aurora VPU	TSMC 16nm FinFET	1.6	~30.22	192	~10.16
A64FX-SVE VPU	TSMC 7nm FinFET	1.8	~1.22	32	~47.2
Vitruvius	GF 22nm FD-SOI	1.4	1.1	16	~20.4

プロセス微細化でさらに高効率化可能と主張

*10 DP-FLOP/cycle: 1サイクルあたりの倍精度(64bit)浮動小数点演算能力

出典：表は2021 RISC-V Summitの発表資料を引用

4-2-3. Advanced Interrupt Architecture (AIA) and Advanced CLINT*11

--- Western Digital (1/2)

(1) 本発表が応用される分野、製品

PLIC*12 (RISC-Vの既存の割込方式) では対応できない分野、製品

(2) 当該分野における技術ニーズ

割込の分散実装、構成可変性、MSI*13対応、仮想化、アドレス空間の効率的な使用など

(3) 本発表の概要

- ・上記ニーズを満たすAIAを定義
- ・SiFive社のCLINTと上位互換のAdvanced CLINT (ACLINT) を定義

(4) 技術的に特筆すべき点

CPUコアベンダー視点で定義された割込仕様の制約を明らかにし、システムベンダー視点で割込を定義

(5) 本発表が目される理由

Linuxをポーティングするために作った基本的な割込仕様 (PLIC) を、リアルタイムOSや複雑な実製品を想定した高度な割込み処理仕様 (AIAとAdvanced CLINT)に拡張し、分散実装、MSI対応などを図った点

(6) 実用化までの課題

- ・ソフトウェアの開発・検証の完遂
- ・Western Digital社以外の第三者のレビューによる仕様の汎用性の確認・確保

*11 Core Local Interrupt

*12 Platform-Level Interrupt Controller

*13 Message Signaled Interrupts

【割り込み処理】

外部からの割り込み要求を受けたとき、プロセッサが割り込みを受け付けられる状態の場合に、実行しているプログラムを一時中断して他のプログラムを実行すること。割り込みしたプログラムが終了後、元のプログラムに戻り、処理を再開する。

4-2-3. Advanced Interrupt Architecture (AIA) and Advanced CLINT*11 --- Western Digital (2/2)

■ 拡張仕様の推奨応用範囲

- **PLIC**がMSI(Message Signaled Interrupts) 非対応だったのに対し、**AIA**はPLICの持っていたすべての制限に対応し、対応の幅を広げている。
- **Advanced CLINT**はタイマーとIPIs(プロセッサ間割り込み)対応のための既存のSiFiveのCLINT仕様の上位互換拡張仕様。既存のRISC-Vプラットフォームとの互換性があり、今後、メジャーな変更はないと考えている。

Recommended ways of using AIA & ACLINT

OS-A Platforms	RISC-V AIA Specification			RISC-V ACLINT Specification			RISC-V SBI Specification			RISC-V Privilege Specification		
	M-level	S-level	VS-level	M-level	S-level	VS-level	M-level	S-level	VS-level	M-level	S-level	VS-level
Legacy Wired IRQs	NA	NA	NA	PLIC	PLIC	PLIC (Emulate)	MSWI (CLINT)	SBI IPI *16	SBI IPI	MTIMER (CLINT) *17	SBI Timer	SBI Timer
Only Wired IRQs	NA	NA	NA	APLIC M-level	APLIC S-level	APLIC S-level (Emulate)	MSWI *18	SSWI *19	SBI IPI	MTIMER	Priv Sstc	Priv Sstc
MSIs and Wired IRQs	IMSI M-file *15	IMSI S-file	IMSI S-file (Emulate)	APLIC M-level	APLIC S-level	APLIC S-level (Emulate)	IMSI M-file	IMSI S-file	SBI IPI	MTIMER	Priv Sstc	Priv Sstc
MSIs, Virtual MSIs and Wired IRQs	IMSI M-file	IMSI S-file	IMSI VS-file	APLIC M-level	APLIC S-level	APLIC S-level (Emulate)	IMSI M-file	IMSI S-file	IMSI VS-file	MTIMER	Priv Sstc	Priv Sstc

*14 IPI : Inter-Processor Interrupt *15 IMSIC : Incoming Message Signaled Interrupt Controller

*16 SBI: Supervisor Binary Interface *17 MTIMER: M-mode Timer *18 MSWI: M-mode Software Interrupts

*19 SSWI: S-mode Software Interrupts

出典 : 表は2021 RISC-V Summitの発表資料を引用

4-3. セキュリティ

4-3-1. SYSTEMATICALLY SECURING THE RISC-V ---- NVIDIA (1/4)

(1) 本発表が応用される分野、製品

Gaming, HPC, Healthcare, Smart City, Robotics, Automotive

(2) 当該分野における技術ニーズ

- ・セキュリティの強化・対策コスト低減
- ・攻撃の高度化（上位レイヤーからシステムの中核レイヤーへ、HW/SWのバグによる脆弱性を攻撃）への対応

(3) 本発表の概要

目的：複雑かつ進化する機能のリスクを安全に管理する基盤を構築し、必須となるセキュリティ機能をあらゆる場所に普及させる

NVRISCV：NVIDIAのRISC-Vコア、Peregrine：NVRISCV + 周辺IP

MILS（Multiple Independent Levels of Security/Safety）方式によるPartitioning

SPARK：信頼性が必須な産業分野で従来から使われている、プログラミング言語 + 分析ツール

BootROMとSeparation Kernelに使用

Offensive Securityによる耐攻撃性評価

ASIL-D*²⁰対応のDual Core Lock Step

(4) 技術的に特筆すべき点

制限のないパーティション構成が可能な柔軟なアーキテクチャ

(5) 本発表が注目される理由

NVIDIA社のセキュリティ対応を詳細に説明

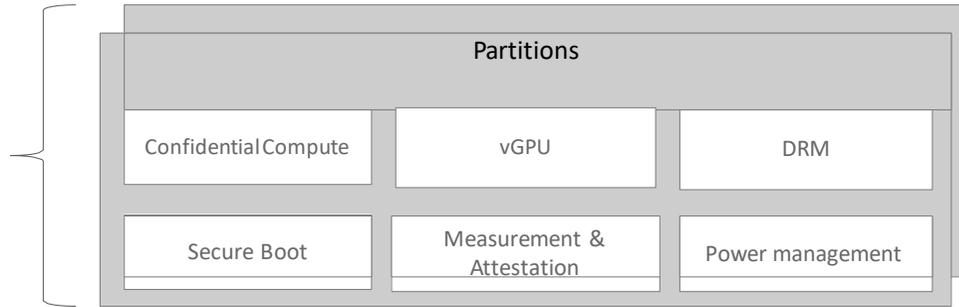
(6) 実用化までの課題

実用化レベル

*20 ASIL(Automotive Safety Integrity Level:自動車安全水準): ISO 26262規格で定義されたリスク分類 Dが最も高い安全性を示す

4-3-1. SYSTEMATICALLY SECURING THE RISC-V --- NVIDIA (2/4)

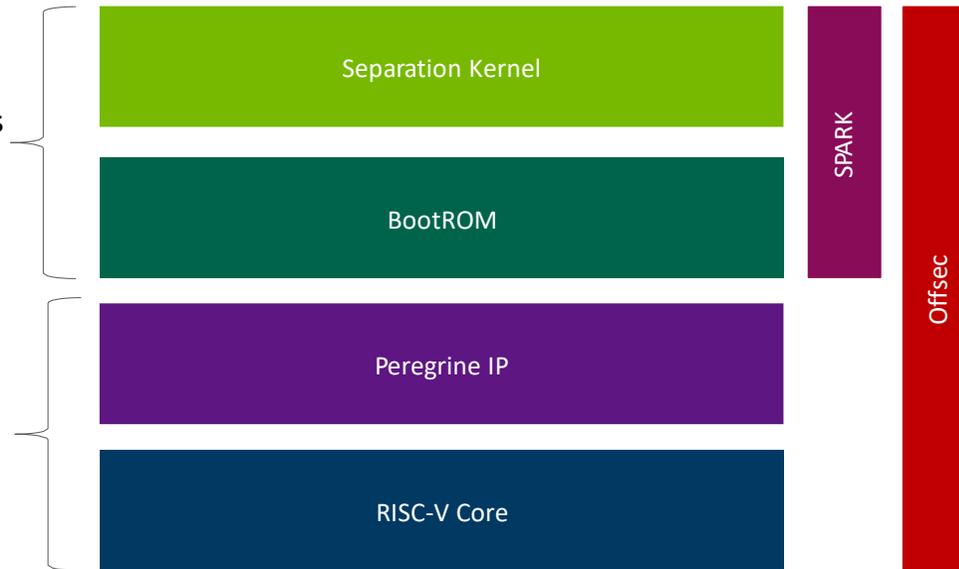
- TBI/Pointer Masking
- Static Analysis
- Fuzz testing



BENEFITS

- Silicon area savings
- Engineering savings
- Security cost savings

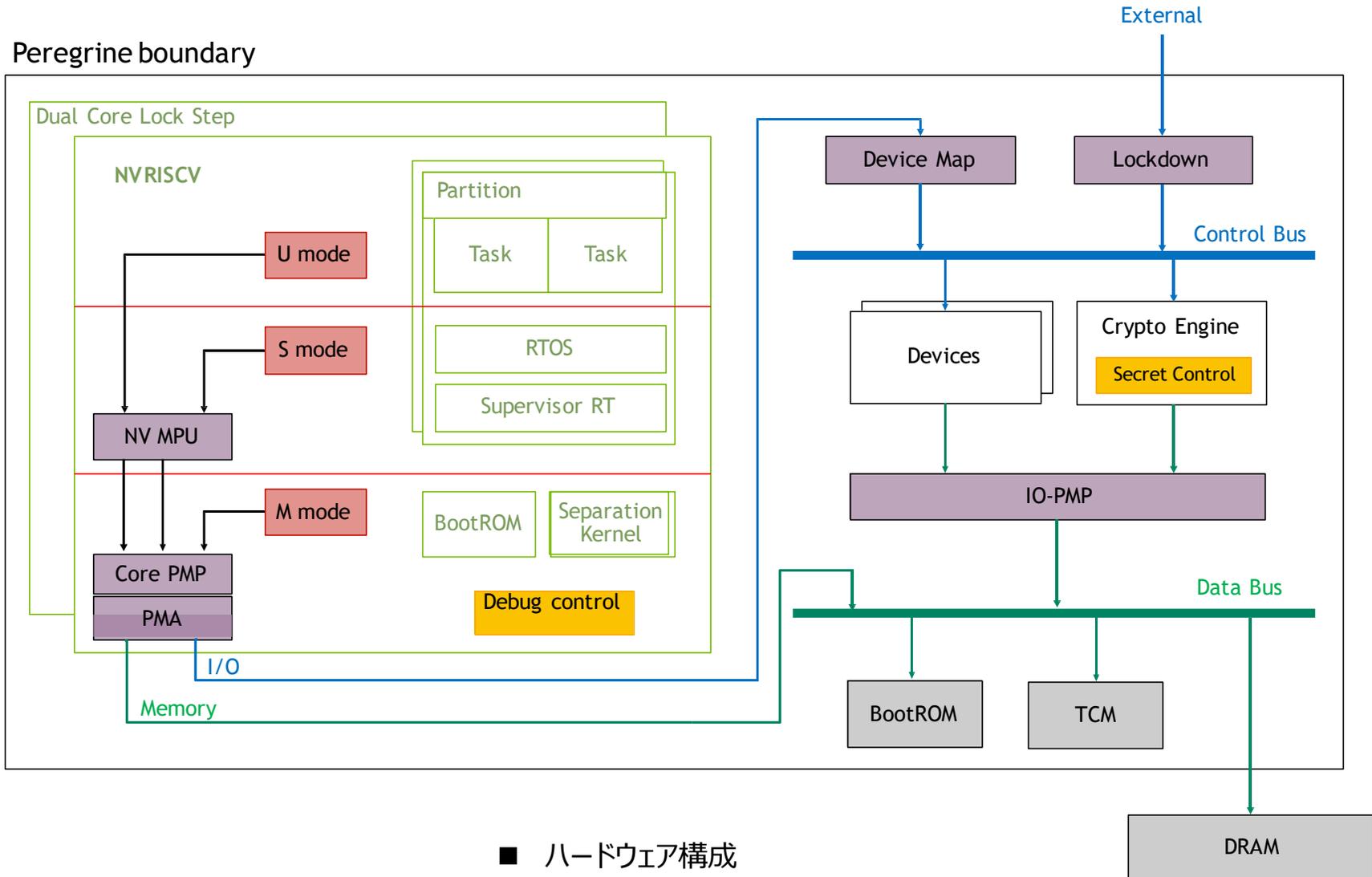
- Formal Verification
- Compiler mitigations
- SW mitigations



- Formal Verification
- Fault Injection Protections

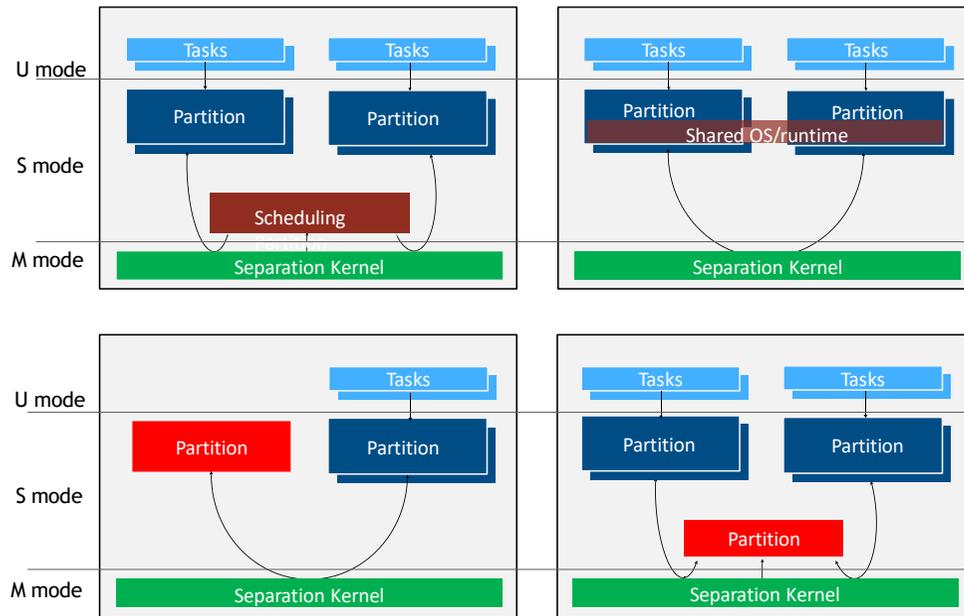
■ 全体構成

4-3-1. SYSTEMATICALLY SECURING THE RISC-V --- NVIDIA (3/4)

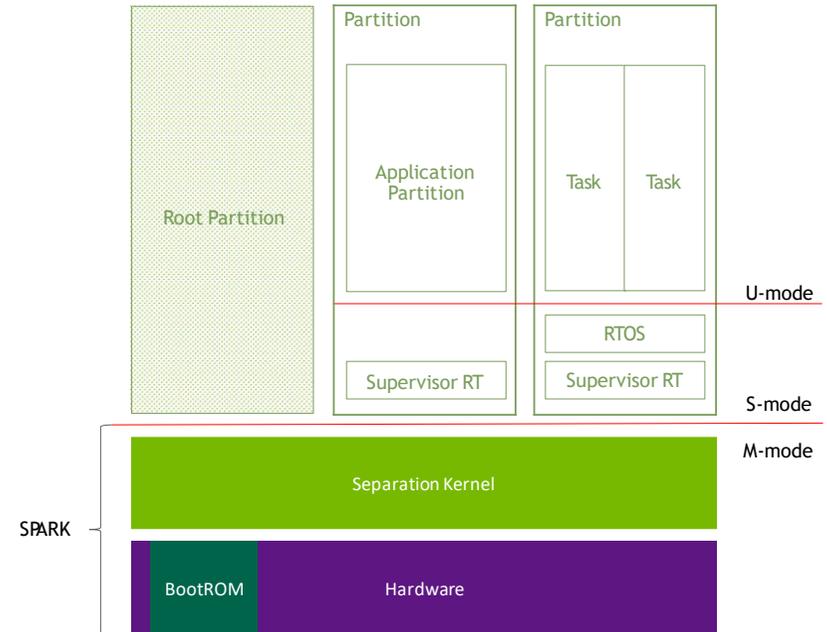


出典：図は2021 RISC-V Summitの発表資料を引用

4-3-1. SYSTEMATICALLY SECURING THE RISC-V --- NVIDIA (4/4)



- 制限のないパーティション構成が可能な柔軟なアーキテクチャ



- Multiple Independent Levels of Security/Safety (MILS) architecture

RISC-Vの特権モード

- M mode: Machine mode
- S mode: Supervisor mode
- U mode: User mode

出典：図は2021 RISC-V Summitの発表資料を引用

4-3-2. The Future of RISC-V Heterogeneous Embedded Virtualization Architectures --- Minho大学 (ポルトガル) (1/3)

(1) 本発表が応用される分野、製品

Flight Controller, Mission Computer, Nano drones, standard sized dronesなど幅広い性能レンジの制御を想定

(2) 当該分野における技術ニーズ

複数サブシステム対応 : Heterogenous Oses (NuttX^{*21}) + Linux), Mixed criticality, Partitioning

(3) 本発表の概要

- Hypervisor拡張を定義し、CVA6 CPUに実装
- CVA6 (Arianeから改称) はOpenHW GroupがメンテしているCPUコア
- ヨーロッパの複数の大学、アブダビの研究機関 (TII), 大学の共同研究成果
- Supervisor mode(S mode)をHypervisor-extended supervisor (HS) とVirtualized supervisor (VS) に分離し、その上にVirtualized User (VU) modeを実装

(4) 技術的に特筆すべき点

- Partitioningとvirtualizationの分離、firmware層でのPartitioning
- RISC-VのPMP拡張^{*22} (ePMP, 2021/12にversion 1.0を正式リリース) を実装

(5) 本発表が注目される理由

誤動作が許されないデバイスの制御をRISC-Vで実現しようとしている点

(6) 実用化までの課題

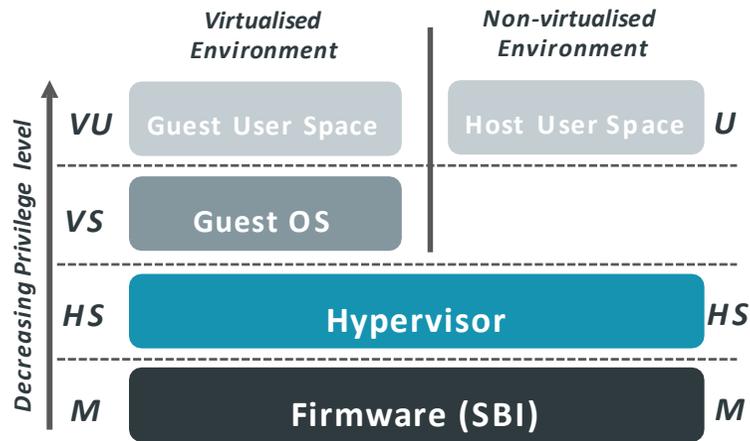
- 定義はほぼ完了しているがHypervisor and Virtual Supervisor CSRs(Control and Status Registers)が未完
- IOMMU, IOPMP(I/O Physical Memory Protection)を活用したPartitioningの定義
- AIA対応

*21 POSIXインタフェースのリアルタイムOS

*22 <https://raw.githubusercontent.com/riscv/riscv-tee/main/Smepmp/Smepmp.pdf>

*23 IOMMU(Input/Output Memory Management Unit): デバイスとメモリのCPUを介さない直接アクセスを制御する

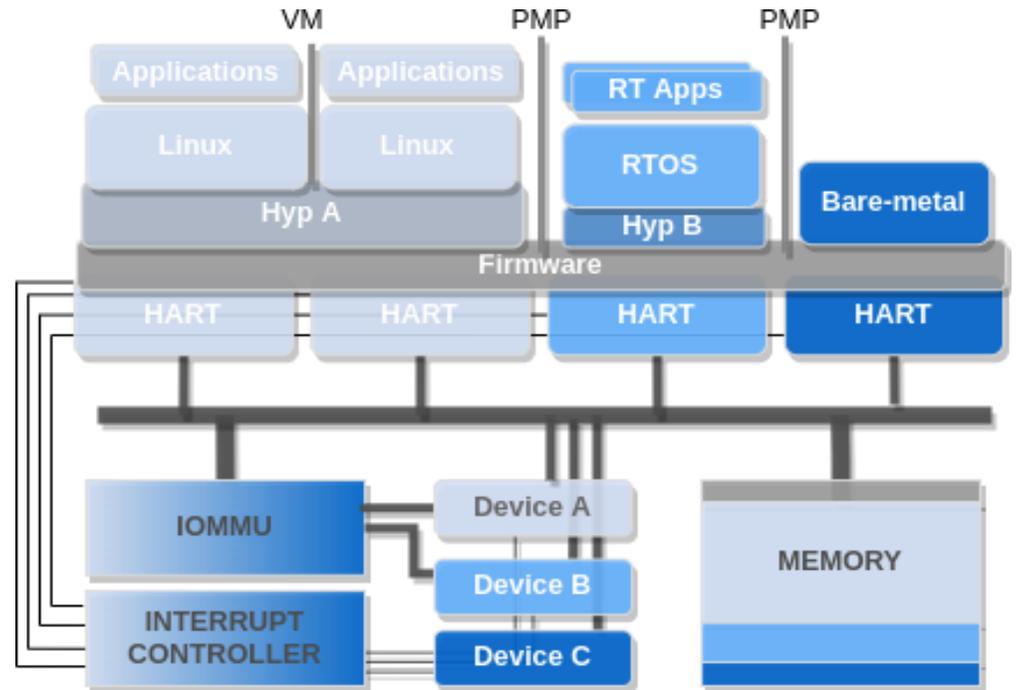
4-3-2. The Future of RISC-V Heterogeneous Embedded Virtualization Architectures --- Minho大学 (ポルトガル) (2/3)



Adapted from: Alistair Francis (WD), "Developing the RISC-V Hypervisor Extensions in QEMU", 2019

■ Hypervisor拡張のイメージ

Supervisor modeを
Hypervisor extended supervisor (HS) と
Virtualized supervisor (VS) に分離し
その上にVirtualized User (VU) modeを実装



■ Partitioningのイメージ

Virtual Machine (VM), PMPによるPartitioning

4-3-2. The Future of RISC-V Heterogeneous Embedded Virtualization Architectures --- Minho大学 (ポルトガル) (3/3)

CVA6のSmepmp拡張の概要

- **RISC-V Smepmp**
 - Protect Machine-mode from less-privileged modes
 - Machine Security Configuration (mseccfg):
 - Rule Locking Bypass (RLB) – bit2
 - Machine Mode Whitelist Policy (MMWP) – bit1
 - Machine Mode Lockdown (MML) – bit0
- **CVA6 Smepmp Implementation**
 - Compliant with version 0.9.5^{*})
 - Minimal changes (11 files, ~100 SLoC)
- **Smepmp Validation & Evaluation**
 - Baremetal testing framework
 - Hardware resources (FPGA) -> 0.1% LUTs

*現在はv1.0をリリース、仕様策定にはNEDO PJ
「セキュアオープンアーキテクチャ基盤技術と
そのAIエッジ応用研究開発」メンバーも参画



PMP Enhancements for memory access and execution prevention on Machine mode (Smepmp)

Nick Kossifidis, Joe Xie, Bill Huffman, Allen Baum, Greg Favor, Tariq Kurd, Fumio Arakawa, RISC-V TEE Task Group

Version 0.9.5, 11/2021: This document is in the Frozen state. Change is extremely unlikely. For more information see: <https://riscv.org/spec-state>

<https://github.com/riscv/riscv-tee/blob/main/Smepmp/Smepmp.pdf>

4-4. Esperanto社のET-SoC-1 (1/4)

- Accelerating AI and non-AI Workloads with 1000+ Energy-Efficient RISC-V Cores on a Single Chip
- Esperanto's Custom RISC-V ISA Extensions for Energy-Efficient Machine Learning Applications

(1) 本発表が応用される分野、製品

- 高並列タスクの実行必要とする分野、製品
- AIをターゲットに開発しているが、非AIにも適用可能と主張

(2) 当該分野における技術ニーズ

- 高並列タスクに適用対象を絞ることで得られる
逐次を含む汎用タスク向けを桁で上回る高効率・高性能化

(3) 本発表の概要

- RISC-Vを1000コア以上集積したチップ
(1088 ET-Minions, 4 ET-Maxions, 1 service processor)
- 高い性能を低電力で実現するために独自のVector/Tensor拡張を定義・実装
- チップについてはCOOL Chips, HOT Chipsなどで発表済み
- 今回は有資格顧客向けの早期アクセスプログラム開始をアナウンス
- また、SamsungのSSDの制御に適用した例を紹介した。

(4) 技術的に特筆すべき点

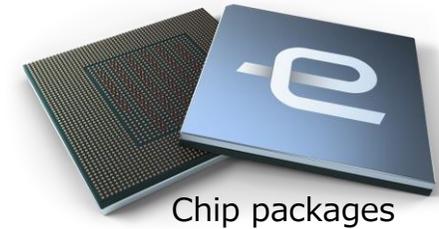
- 非常に高い性能/電力が特長、**低電圧 (0.3-0.4V) 動作**が貢献

(5) 本発表が注目される理由

- ベンダー独自の命令拡張を許容しているRISC-Vの特徴を生かして、**汎用コアをアクセラータ化し高並列化することで、汎用コアが担うメモリアクセスが従来方式よりボトルネックになりにくいこと**

(6) 実用化までの課題

- 実用化レベル



Chip packages



Chip layout

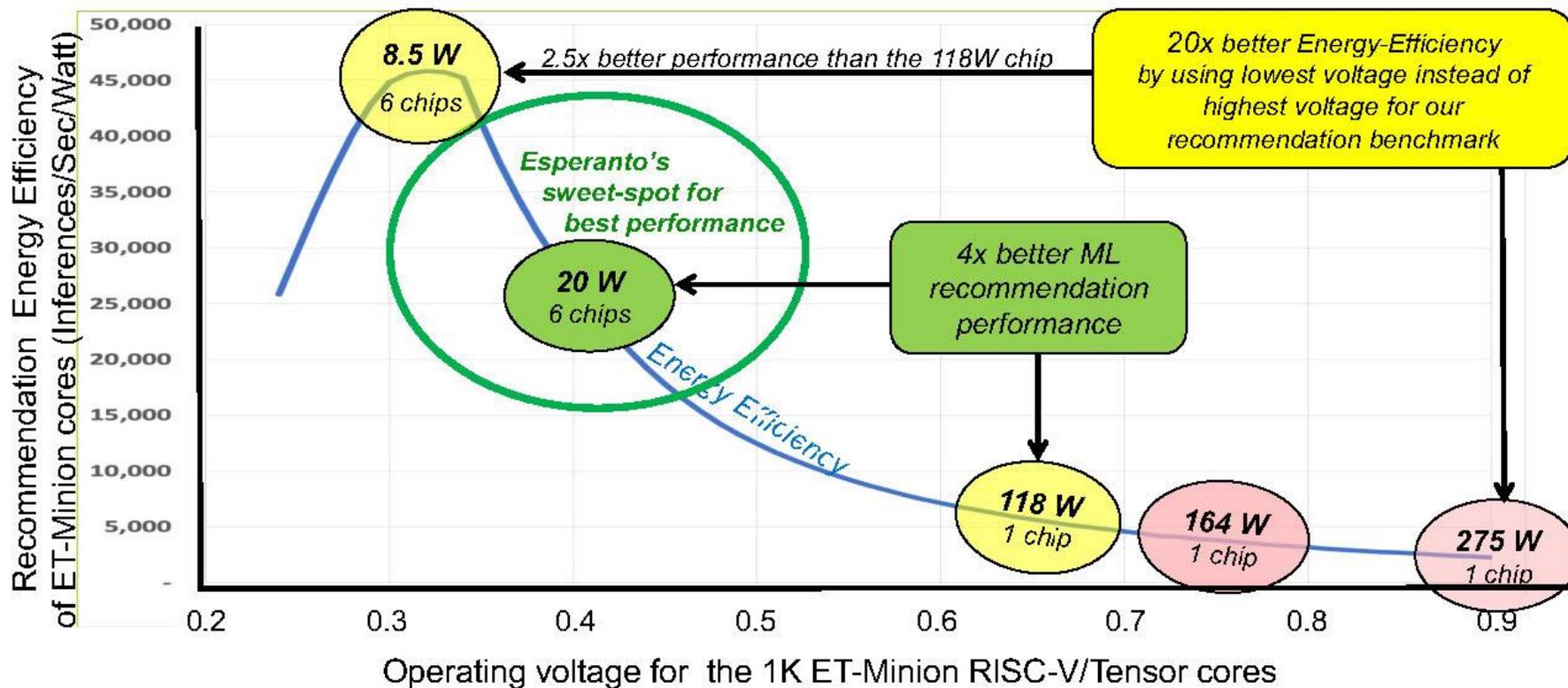


6 chip board

出典：図は2021 RISC-V Summitの発表資料を引用

4-4. Esperanto社のET-SoC-1 (2/4)

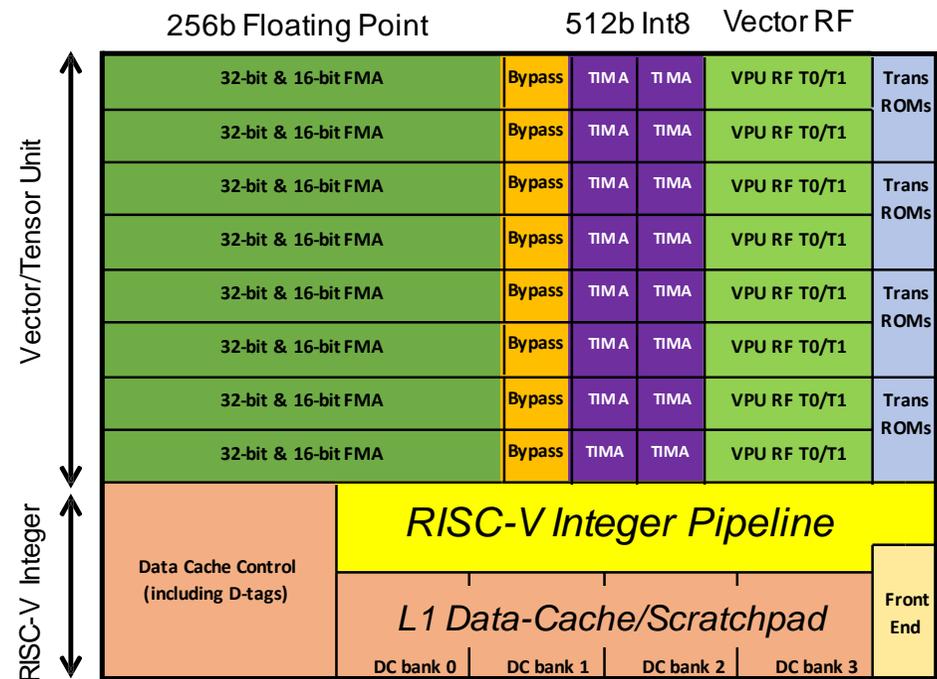
- 低電圧動作による電力効率向上
- 効率と性能のバランスから6チップボードを0.4V、20Wで動作させるのが最適



出典：図は2021 RISC-V Summitの発表資料を引用

4-4. Esperanto社のET-SoC-1 (3/4)

- ET-Minion : 独自のVector/Tensor Unitを備えたエネルギー効率の高い64ビットRISC-V CPU
 - インオーダーパイプライン、2つのハードウェア実行スレッド
 - 低電圧動作に最適化されたアーキテクチャと回路
 - 低ゲート段数/ステージにより低電圧でも高MHz
 - ソフトウェアで構成可能なL1データキャッシュ/スクラッチパッド
- MLに最適化されたVector/Tensor Unit
 - 512b Int8
128並列の8ビット整数処理、累算は32ビット
 - 256b Floating Point
16並列単精度、32並列半精度
- 複数サイクル実行のTensor命令
 - 最大で32K処理/512サイクルを1命令で実行
 - 命令フェッチのバンド幅と電力を低減
 - その間Integer Pipelineはスリープ
- Vector超越関数命令
- 300MHz ~ 2GHzで動作



ET-Minion RISC-V Core and Tensor/Vector unit optimized for low-voltage operation to improve energy-efficiency

4-5. SiFive (1/9)

- Keynote: The Future of RISC-V Has No Limits (Dr. Yunsup Lee Co-Founder and CTO)
- Implementing Functionally-safe RISC-V IP for Automotive and Safety Critical Applications (Shubu Mukherjee, VP Architecture)
- Continuous Innovation in Embedded RISC-V Processors (Drew Barbier Senior Director of Product Marketing)

(1) 本発表が応用される分野、製品

RISC-VプロセッサIPコアの提供、自動車、モバイルエッジ処理からデータセンターコンピューティングの幅広い分野をターゲット

(2) 当該分野における技術ニーズ

市場はRISC-Vを必要としているが性能は、幅広い採用のための障害となる。SiFiveはエコシステムの中でより多くのプロセッサIPを提供し、市場投入時間を短縮する。

(3) 本発表の概要

- Performance P650プロセッサは、Performance P550プロセッサのわずか数か月後に登場し、**前モデルを50%上回っている。**
- 自動車分野への適用は、高い性能とセキュリティが求められ、またゾーンアーキテクチャへの進化が求められている。
(4-5.の(6/9)自動車分野への適用アーキテクチャ、(9/9)SiFiveでのセキュリティ機能の実装で詳細解説)

(4) 技術的に特筆すべき点

コヒーレントマルチコアコンプレックス*24を使用して16コアに拡張できる。これには、メモリ管理から割り込みユニット制御に至るまでのシステムコンポーネントが含まれ、仮想化に必要なRISC-Vハイパーバイザー拡張が可能。

(5) 本発表が注目される理由

- SiFiveは2015年のRISC-V立ち上げ時期以降、エコシステムのリーダ的役割を担っている。
- Intelが2022年に立ち上げる7nmラインの開発プラットフォームにSiFive P550を採用した。**
- ルネサスとSiFiveは、**自動車用途向けのハイエンドRISC-Vソリューションを共同でリリースした。**

(6) 実用化までの課題

特になし

*24 マルチコアシステムの課題であるデータの整合性を担保するための仕組み

4-5. SiFive (2/9)

4-3-1. SYSTEMATICALLY SECURING THE RISC-V --- NVIDIA (1/4)

■ SiFiveのパートナー事例

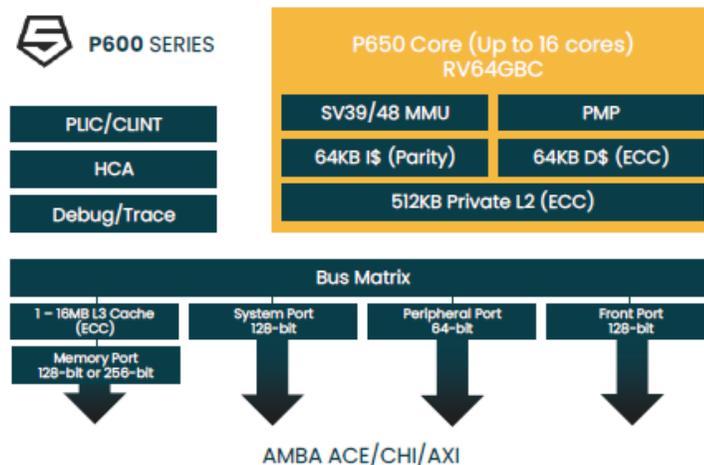
No	会社名	内容
1	Intel	2022年に立ち上げる7nmラインの開発プラットフォームにSiFive P550を採用
2	ルネサス	自動車用途向けのハイエンドRISC-Vソリューションを共同でリリース
3	Tenstorrent	有名なチップアーキテクトJim Keller氏の元、AIチップGraySkullに512ビット幅のRISC-Vベクトル拡張(RVV)を統合したX280 64ビットRISC-Vベースのコアを採用
4	Samsung	Samsung第2世代14LPP FinFET技術を用いたAI/MLカスタムSoCのプラットフォームを顧客に提供開始
5	Canonical	・SiFive FU740 SoCが搭載しLinuxが動作するHiFive Unmatched 開発ボードをSiFiveが2021年6月に販売※ ・CanonicalがRISC-VにUbuntuを移植

※電子部品の調達難により、2022年1月、在庫限りでの販売終了をアナウンス

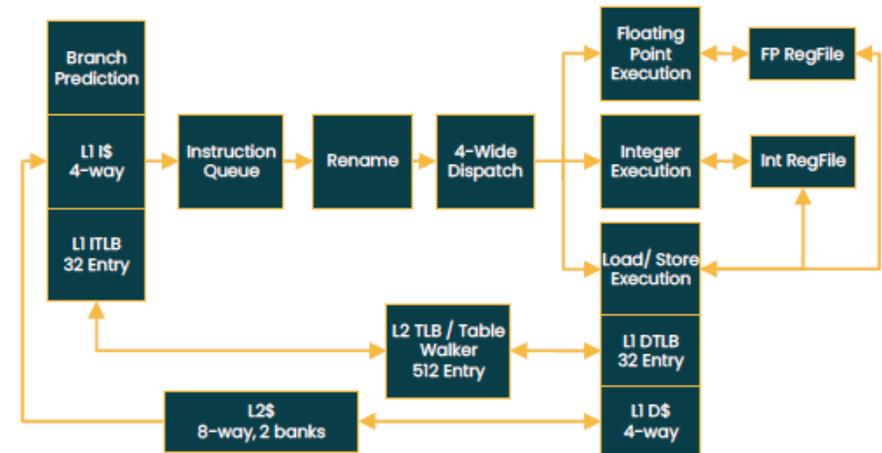
4-5. SiFive (3/9)

■ Performance P650プロセッサの概要

- Quad-Issue
- Out of Order
- 64ビットISAは倍精度FPUとビット操作拡張機能を含む
- コアは64KB命令キャッシュ、64KBデータキャッシュ、512KB L2キャッシュ
- マルチクラスターコヒーレンスをサポートし、最大4つのクラスター、16コアの複合コアに拡張できる
- 性能：11+ SPECInt2k6/GHz 2.75GHz+ in 5nm
- 高度なセキュリティ機能と暗号アクセラレータを持つ
- IPとして、事前に統合および検証済のVerilog RTLを提供可能



ブロックダイアグラム^{*25}



コアアーキテクチャ

*25 AMBA(Advanced Microcontroller Bus Architecture): SoCを構成するために必要なインターフェイス等の規格 ACE(AXI Coherency Extensions)、CHI(Coherent Hub Interconnect)、AXI(Advance eXtensible Interface)等がある。

出典：図は2021 RISC-V Summitの発表資料を引用

4-5. SiFive (4/9)

■ Performance P650プロセッサのP550との比較

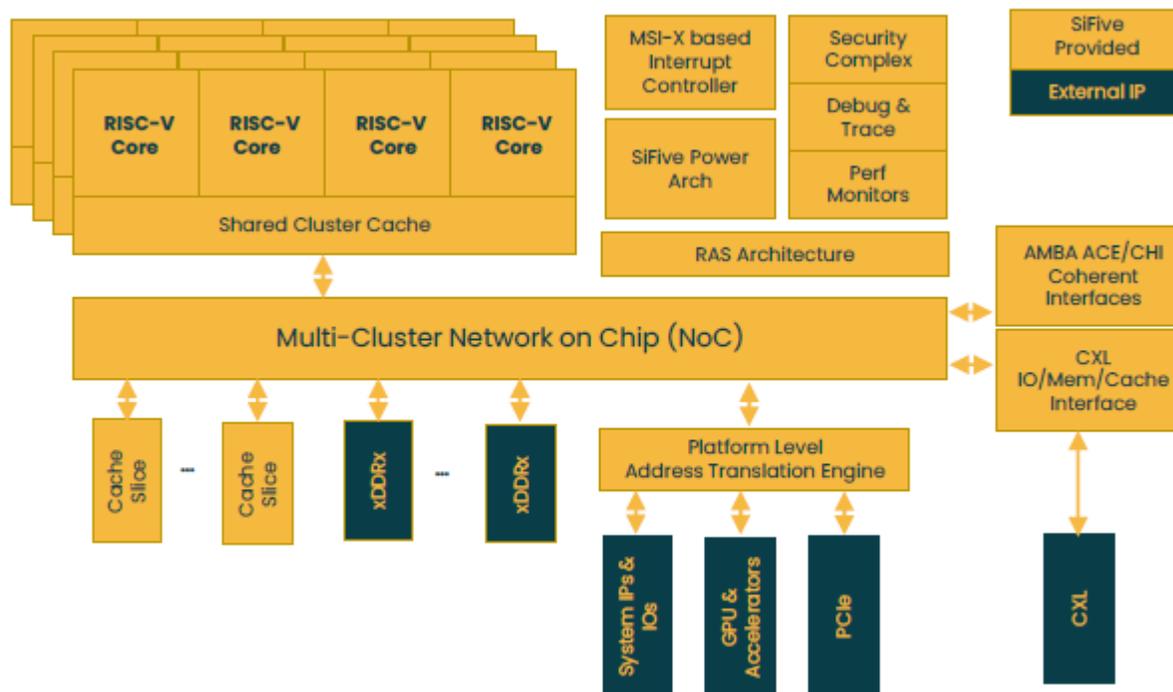
Design Parameter	SiFive Performance™ P550	SiFive Performance™ P650
Decode	Up to 3 instructions per cycle	Up to 4 instructions per cycle
Integer ALUs	3; 1x Branch, 1x Mul/Div	4; 2x Branch, 1x Mul/Div
Floating Point Pipelines	2; 1x Divide	2; 1x Divide
Load/Store Units	1 Load Unit, 1 Store Unit	2 Generic Load Store Units
Integer Registers	100%	162%
Floating Point Registers	100%	173%

出典：表は2021 RISC-V Summitの発表資料を引用

4-5. SiFive (5/9)

■ Performance P650プロセッサのシステムソリューション（パートナーと開発中）

- 複合コアコンプレックス/クラスター/コアレベルの電力監視と管理
- コヒーレントアクセスポイントを備えたスケーラブルなクラスタとNoC
- AMBA ACE / CHI, CXL
- 高帯域幅およびスループットメモリサブシステム
- コアおよびシステムレベルで有効化された仮想化
- FuSa(Functional Safety : 機能安全)対応の高度なエラー検出、レポート、およびリカバリ
- 高度なパフォーマンスの監視、デバッグ、およびトレース

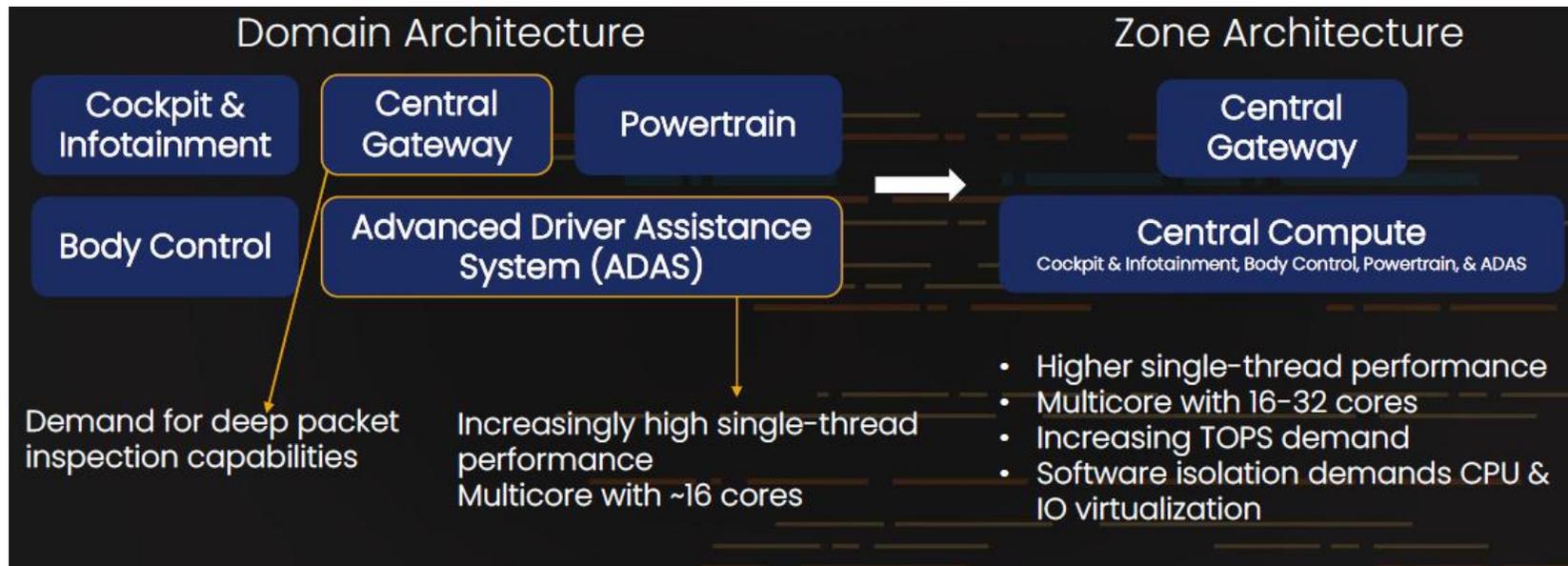


出典：図は2021 RISC-V Summitの発表資料を引用

4-5. SiFive (6/9)

自動車分野への適用アーキテクチャ

- Implementing Functionally-safe RISC-V IP for Automotive and Safety Critical Applications (Shubu Mukherjee, VP Architecture)
- 自動車分野では、ドメインアーキテクチャーからゾーンアーキテクチャーへの進化が求められている。
- ドメインアーキテクチャはHW Isolation、ゾーンアーキテクチャはSW Isolationというのが最も大きな違いで、仮想化によって、HW Isolationと同じ強度のIsolationが可能になり、移行が実現。制御対象の数だけ独立な制御HWが必要だったものが、統合で数を減らせ、さらに連携が容易になるというメリットがある。
- 自動運転は、高い性能とソフトウェアの高い独立性が求められる。
 - 16-32コアが必要
 - 単体のスレッド自身は高性能が要求され、トータルの要求TOPSは大きい（* 最近のGrayskaull, Tesla DOJO, Mobileye EyeQ Ultraの発表を見ると数100TOSレベルが必要であり、RISC-Vコア+GPU/AIアクセラレータで実現している）
 - セキュリティを考慮するとIO、CPUの仮想化が必要（SiFive (9) SiFiveでのセキュリティ機能の実装を参照）

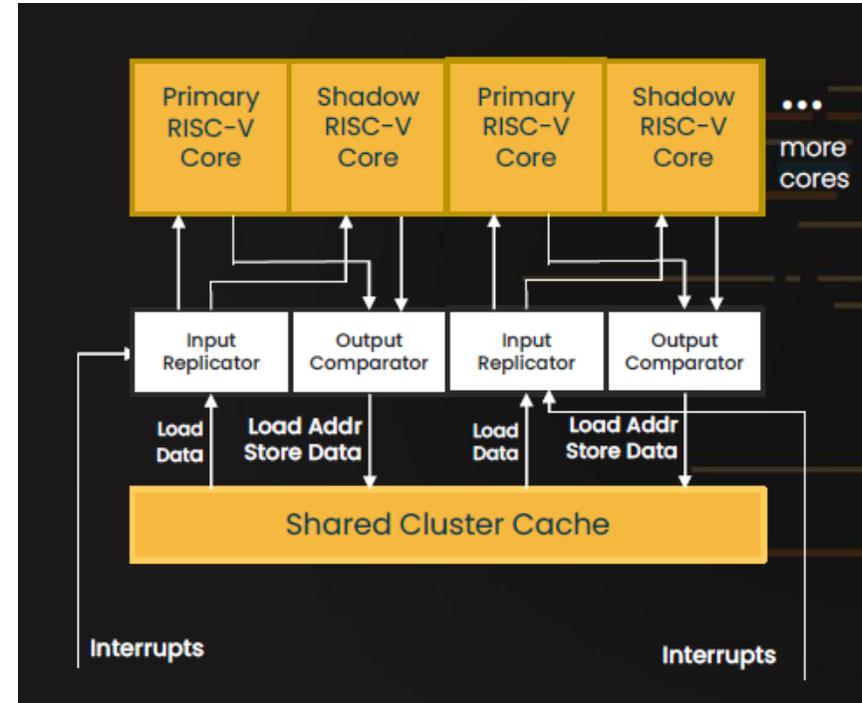


出典：図は2021 RISC-V Summitの発表資料を引用

4-5. SiFive (7/9)

自動車分野への適用アーキテクチャ詳細

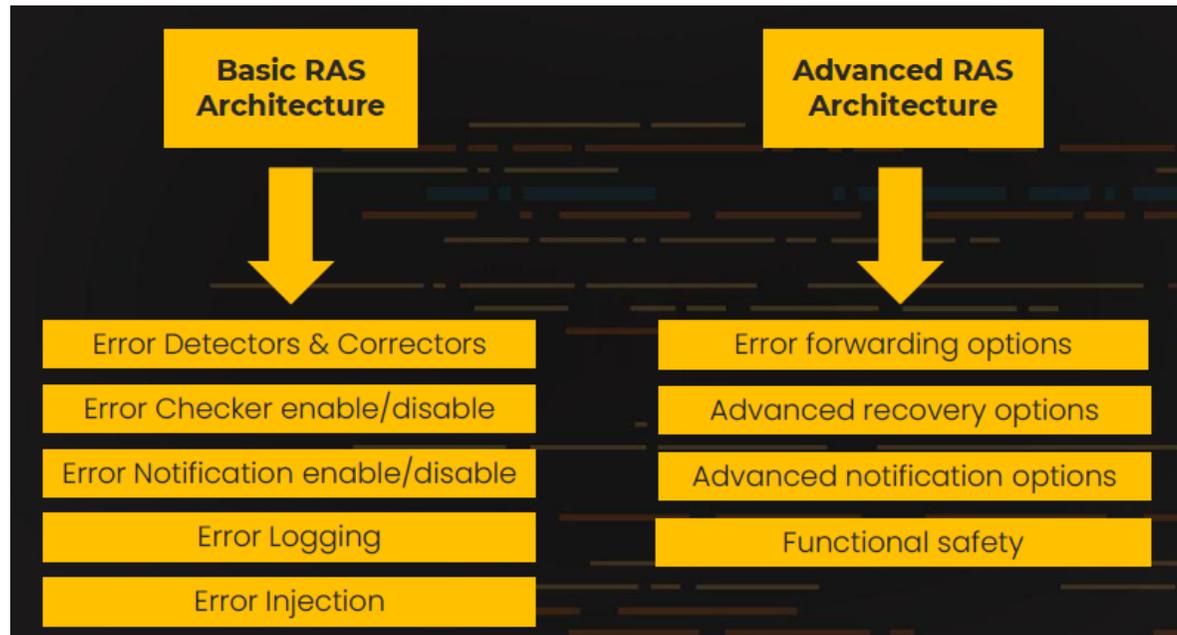
- Dual Core Lockstep アーキテクチャー
 - DCLSEモード
 - ロックステップのコアのペア (プライマリとシャドウ)
 - N個のプライマリアコア+ N個のシャドウコア
 - 分割モード (Split mode)
 - 合計Nコア
 - コンパレータとレプリケータが無効になっている
 - 複数モードのサポート
 - すべてのコアがDCLSEモードで動作
 - すべてのコアが分割モードで動作
 - 一部は分割モード、一部はDCLSEモードで動作 (Split-Lock mode)



4-5. SiFive (8/9)

自動車分野への適用（機能安全）

- Advanced RASアーキテクチャー
 - 機能安全（Functional Safety）をサポート
 - 機能安全とは？
 - 自動車の機能安全規格「ISO 26262:2011」は、IEC 61508の分野規格であり、IEC 61508の基本的考えと枠組みを踏襲している。しかし、全体システムの定量的リスク評価は困難なため、自動車の機能安全規格「ISO 26262:2011」の機能安全の定義は、IEC 61508の定義より狭く、「電気電子(E/E)システムの機能不全のふるまいにより引き起こされるハザードが原因となる、不合理なリスクの不在」[2]となっている。つまり、機能不全のふるまい(誤動作)を対象にしている。機能不全のふるまい(誤動作)が発生しない場合でも、安全な状態ではない場合がある。豪雨などの運転状況下で衝突予防機構によって横滑りや追突が起きることはある。このように、ISO 26262:2011では、豪雨などの天候の影響や、交通法規を著しく逸脱するような運転手の危険運転を監視装置や防護装置によって防ぐという観点がない。



出典：図は2021 RISC-V Summitの発表資料を引用

4-5. SiFive (9/9)

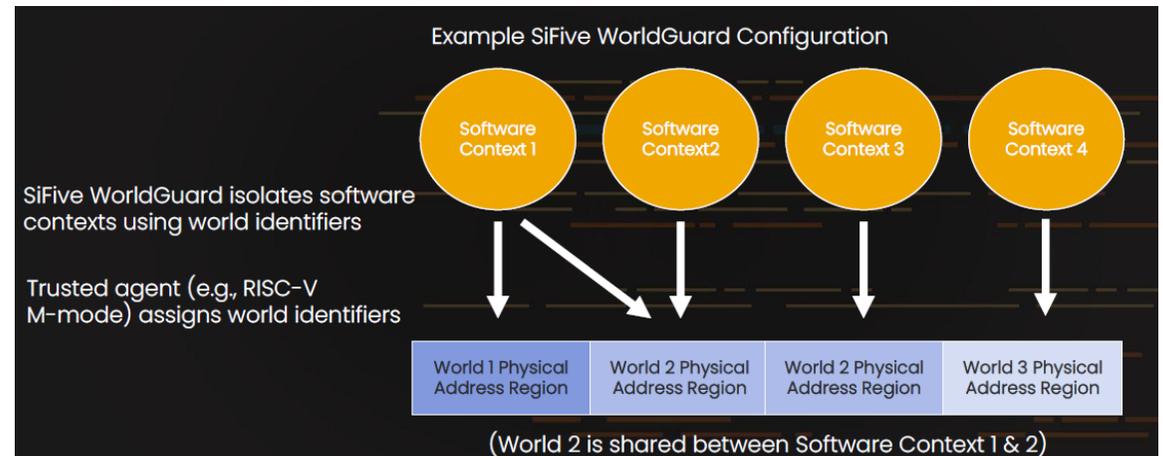
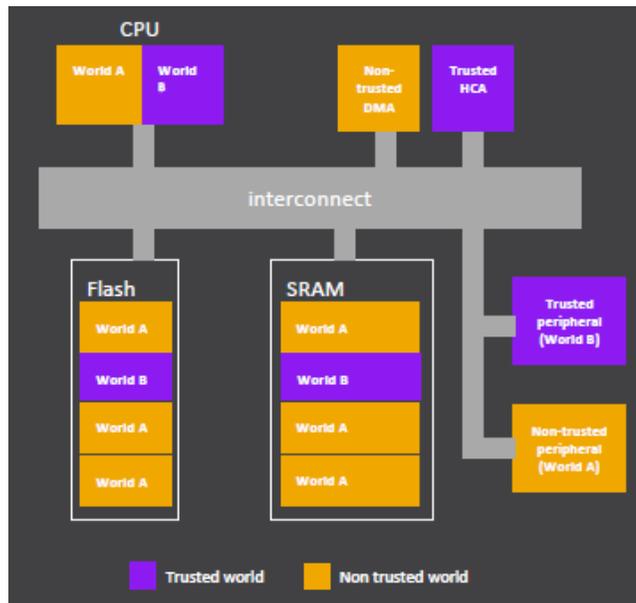
SiFiveでのセキュリティ機能の実装

■ SiFive WorldGuard

- Armは、Arm9の特徴の一つであるArm Confidential Computing Architecture (CCA) で、管理権とアクセス権を分離しユーザーデータを管理者からも保護することで、セキュリティを強化している。そのため、演算処理の行われるあらゆる場所でデータとコードを保護する。
- SiFive WordGuardも、world IDによるSWコンテキスト分離によってデータとコードを保護し、セキュリティを強化している。SoCの中で、メモリ、キャッシュ、CPU、IOをそれぞれTrusted world とNon Trusted worldに分離することが可能である。そして、SWコンテキスト分離のために、仮想化機能を完備している。

■ 暗号エンジン

- インハウスの暗号エンジン（オープンソース）を実装。なお、最近RISC-V ISAにはScalar Cryptography Extensionが定義され、これを実装すればSWによる効率的な暗号エンジン実装も可能である。



出典：図は2021 RISC-V Summitの発表資料を引用

4-6. Andes Technology社 (1/3)

- Keynote: Beefing Up the Datacenter Accelerators
- IOPMP Updates: The Protection of IOPMP

(1) 本発表が応用される分野、製品

1) Mobile, 2) AIoT & Auto, 3) Storage & 5G, 4) Datacenter

(2) 当該分野における技術ニーズ

1) MCU制御, 2) アプリ制御とDSPによる高速化, 3) Manycoreと逐次処理, 4) Manycoreとvector処理

(3) 本発表の概要

- 製品ラインナップと主要コアの概要を紹介
- IOPMPの必要性と現状を紹介

(4) 技術的に特筆すべき点

- Andes Custom Extensions (ACE) を定義し、独自のHardwired Engine (HWE) を密結合で活用
- IOPMPの実装 (Andes社はRISC-V InternationalのTEE TGでIOPMP仕様定義を積極的に推進している)

(5) 本発表が注目される理由

- NX27VはRISC-VのVector Extensionを実装した最初の商用コア (昨年発表済み)
- RISC-Vコアの主要ベンダーの一つで、幅広い製品ラインナップがあり、多くのチップベンダーに供給
- RISC-V InternationalのFounding Premium Membersの一社

(6) 実用化までの課題

- 既に製品コアライセンスを提供中
- IOPMPはTEE TGで仕様を固めているフェーズ

4-6. Andes Technology社 (2/3)

- 各分野の主要ベンダーにCPUコアを提供

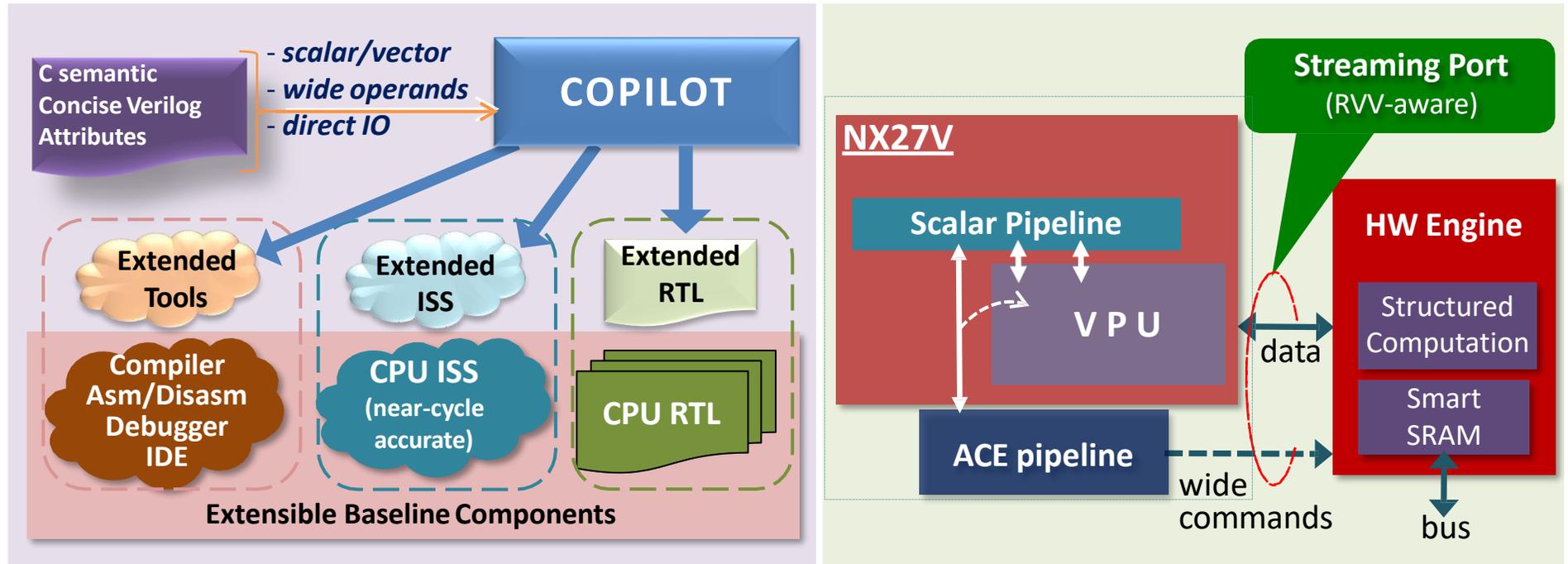
Mobile	AIoT & Auto	Storage	5G	Datacenter
				
<p>MEDIATEK</p> <p>Mobile AP</p> <p>TDDI</p> <p>Billion shipments of WiFi/BT, touch, and sensor hub.</p> <p>MCU control: AndesCore N25F, N45</p>	<p>RENESAS</p> <p>Performance, safety and customization</p> <p>Kneron</p> <p>Telink HPMicro</p> <p>Mainstream AIoT, audio, video, and wearables</p> <p>Application control and DSP acceleration: AndesCore D25F, D45, AX25MP, AX45MP</p>	<p>PHISON</p> <p>Efficient cores for scalable performance</p>  <p>Flexible interfaces for efficient SoC</p> <p>Manycore scalar processing: AndesCore N25F, NX45, AX45MP</p>	<p>EDGE</p> <p>Customizable core for disruptive innovation</p> <p>PICOCOM</p> <p>Manycores for max flexibility and scalability</p> <p>Manycore scalar processing: AndesCore N25F, A25</p>	<p>SK telecom</p> <p>Configurable & powerful cores for Cloud AI</p> <p>LUMINOUS 希姆计算</p> <p>Cloud Solution</p> <p>Vector and ACE-enabled cores for AI Acceleration</p> <p>Manycore vector processing: AndesCore NX27V (+ NX25F, AX27, AX45MP)</p>

出典：図は2021 RISC-V Summitの発表資料を引用

4-6. Andes Technology社 (3/3)

Andes Custom Extensions (ACE)

- RISC-Vは独自命令拡張可能なアーキテクチャ
- Andes社もACEによって命令拡張による性能・効率向上を達成すべく環境を整備している
- ACEはカスタマイズされたレジスタ/メモリ/ポートを使用する新命令作成を容易にする
- さらに、演算命令の設計と、Streaming PortによるHardwired engine (HWE) との密結合を可能にする



出典：図は2021 RISC-V Summitの発表資料を引用

4-7. Micronchip PolarFire® SoC FPGA (1/3)

- Keynote: Microchip and the Expanding RISC-V Universe - Ted Speers, Technical Fellow
- Hard Real-Time vs High Performance Real-Time Applications on PolarFire SoC - Hugh Breslin

(1) 本発表が応用される分野、製品

通信、防衛、産業用制御、組込機器、IoT、車載

(2) 当該分野における技術ニーズ

- 小～中量生産品
- 低消費電力

(3) 本発表の概要

- PolarFire SoC FPGAの概要を紹介
- Mi-V(Microchip's RISC-V initiatives)によるエコシステムの提供

(4) 技術的に特筆すべき点

- RISC-Vを組み込んだフラッシュメモリを使った不揮発性FPGA(SRAMベースのFPGAよりも50%低消費電力)
- RISC-VをハードマクロとしてFPGAに組み込み、柔軟なSoC構成が可能

(5) 本発表が注目される理由

- 2020年には、マイクロコントローラの市場で世界第3位となり、トップのルネサスと3.6%差まで上昇した。
- NVMを使ったFPGAベンダーとしてミッドレンジのFPGA市場で存在感を示しており、エコシステムの作り方などの戦略も参考になる。

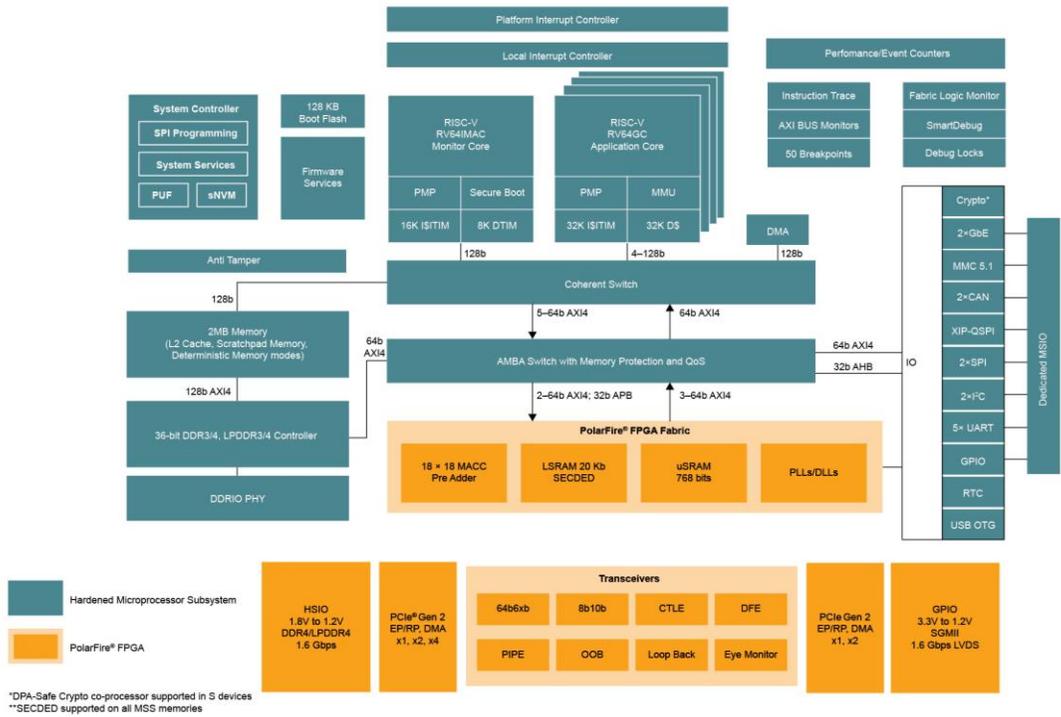
(6) 実用化までの課題

特になし

4-7. Micronchip PolarFire® SoC FPGA (2/3)

■ ブロックダイアグラム

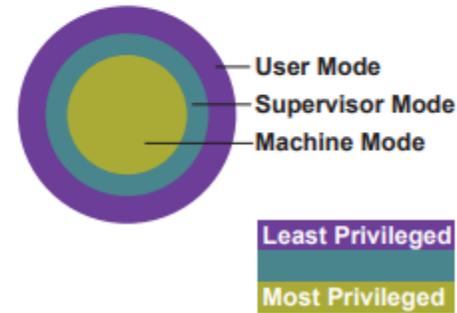
PolarFire® SoC Block Diagram



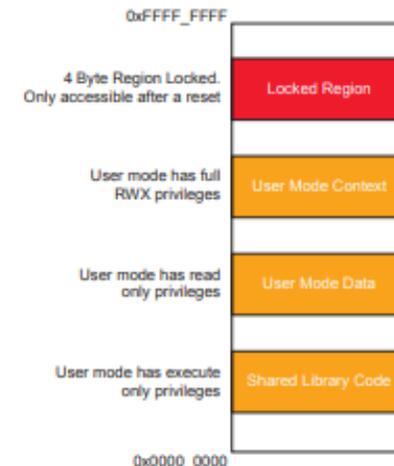
■ セキュリティ

PolarFire SoC の物理的なメモリ保護

PolarFire SoC は各プロセッサコアに物理的なメモリ保護 (PMP) を実装しています。PMP は、優先度が低いモードで (読み書き、実行) の制約を適用するために使います。PolarFire SoC は、信頼できないユーザーモード ソフトウェアのアクセス権を制限できます。



PolarFire SoCでの PMP の実装



出典：図は2021 RISC-V Summitの発表資料を引用

4-7. Micronchip PolarFire® SoC FPGA (3/3)

- Mi-V(Microchip's RISC-V initiatives)によるエコシステムは、PolarFire SoC に合わせて構築された RISC-V エコシステムの一部

オペレーティング システム



RTOS



コンパイラ



ミドルウェア



Mi-V デザイン パートナーおよび SoM ベンダー



4-8. EdgeQ (1/3)

Extending RISC-V Instructions for 5G Intelligent RAN Base Stations - Gururaj Padaki & Sriram Rajagopal

(1) 本発表が応用される分野、製品

5G

(2) 当該分野における技術ニーズ

- ・大手通信機器ベンダーの囲い込みを避け、オープン仕様に基づいて自由にさまざまなベンダーの機器を組み合わせられるようにする「Open RAN」
- ・無線アクセス（RAN）の管理・制御の自律化、AI化による効率化

(3) 本発表の概要

EdgeQのRISC-Vベース5G基地局オンチップのアーキテクチャを紹介

(4) 技術的に特筆すべき点

- ・ソフトウェア+ハードウェアで5Gベースバンド処理を実現
- ・標準のRISC-Vコアに5GおよびAI（人工知能）のワークロードに特化した拡張命令を追加してカスタマイズ

(5) 本発表が注目される理由

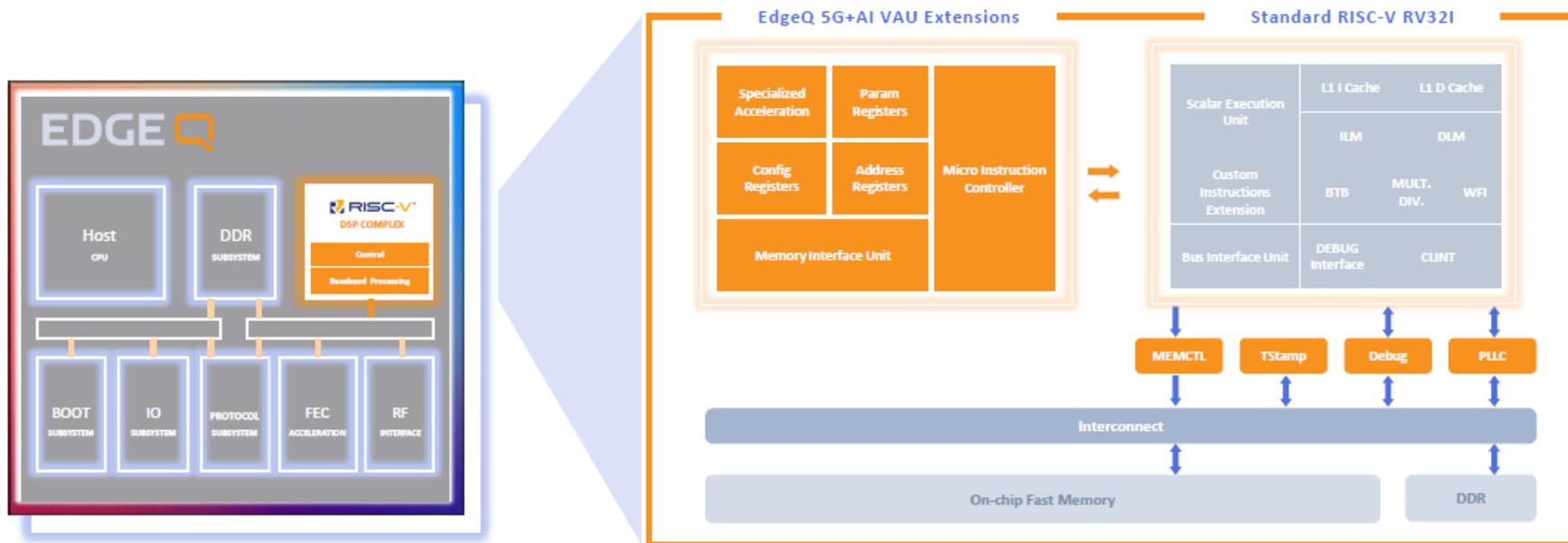
- ・EdgeQはQualcommにルーツを持つスタートアップで、2020年にシリーズAで約4000万ドルを調達した。
- ・基地局を構成するさまざまなチップを1つのプログラマブルチップに統合して、ファームウェアで分散ユニット（DU）、無線ユニット（RU）、アクセスポイント（gNodeB）に変更可能であり、全体の消費電力と基地局を構築するコストを削減する狙いがあり、注目されている。

(6) 実用化までの課題

- ・2021年9月にサンプル出荷
- ・既に、北米の大手OEMのデザインウィンを獲得しているとのこと

4-8. EdgeQ (2/3)

- 標準のRISC-VコアRV321に加え、5GとAI制御のためのVAU (Vector Arithmetic Unit) とVAUを使った信号処理拡張命令セットを持つ。
- 50コアのRISC-Vコアを搭載ソフトウェアで実装されているPHY層の機能を処理するために動的にプログラムできる。同コアは、Andes Technologyからライセンス供与されたもので、5GおよびAI (人工知能) のワークロードに特化したEdgeQの命令拡張でカスタマイズされている。
- EdgeQが追加した命令は、高速フーリエ変換 (FFT) や複素変調、非線形関数の近似器、行列分解、等化などの数学演算をするもので5GとAI処理両方で使われる。



EdgeQ RISC-V with Baseband Extensions

- FFT (Fast Fourier Transform)
- Complex modulations
- Non-linear functions
- Matrix Decompositions
- Equalization
- Function Specific Acceleration
- Multi-support for 4G/LTE and 5G/NR
- More ...



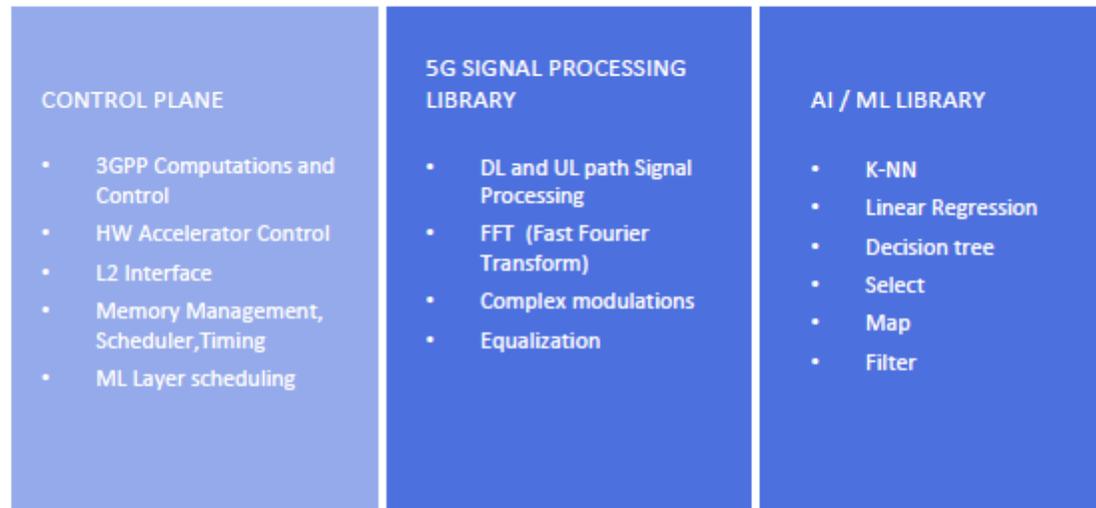
出典：図は2021 RISC-V Summitの発表資料を引用

4-8. EdgeQ (3/3)

- ライブラリはFFTなどの信号処理や、K近傍、回帰、決定着などのAI/MLライブラリをサポート
- 通信もハードで実現するだけでなくソフトで実現、しかも制御はAI的な手法を使って最適化する時代となった。



RISC-V DSP Complex



出典：図は2021 RISC-V Summitの発表資料を引用

4-9. Mythic & Cudasip (1/2)

Lightning Talk: Using and Extending RISC-V in an Analog Matrix Processor for Neural Networks - David Luo, Mythic & Dr Zdeněk Přikryl, Cudasip

(1) 本発表が応用される分野、製品

AIエッジデバイス (AIoT)、応用分野としてはドローン、ロボット、監視カメラ、車載

(2) 当該分野における技術ニーズ

低消費電力

(3) 本発表の概要

- MythicのAIエッジ推論チップのアーキテクチャー紹介
- CudasipがSIMDサポートのRISC-Vコアを提供

(4) 技術的に特筆すべき点

- RISC-V、SIMDエンジン、アナログ計算エンジン、NoC(Network-on-chip) を搭載したAI推論用のチップ
- Deep Learningの重みは不揮発性のフラッシュメモリに格納し、アレイでMAC演算する

(5) 本発表が注目される理由

エッジでの低消費電力を実現するインメモリコンピューティング手法の一つだが、0.5pJ/MACと低消費電力

(6) 実用化までの課題

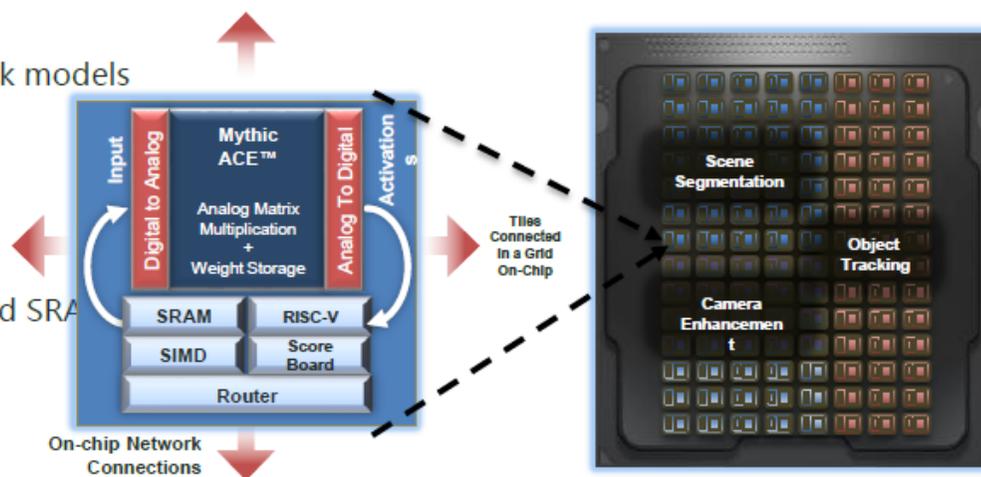
- 試作完。2018年のHotchip30で発表済
- 量産ベースにするには、アナログは製造ばらつき、温度、フラッシュメモリのVthシフトなどがあり補正回路を組み込む必要がある。

4-9. Mythic & Codasip (2/2)

- アナログNVM+アナログ乗算+DAC/ADC+デジタルRISC-Vが一つのTile。一つのチップに76個のTile。
- 1つのチップのタイルグリッドを分割して、シーンのセグメント分割、オブジェクトの追跡、カメラ画像のエンハンスなど、異なる処理を行わせることもできる。

Architecture of Mythic Analog Matrix Processor (AMP™)

- Mythic AMP integrates a large array of tiles
 - Support for multiple concurrent Neural Network models
 - All weights on-chip
- Each Mythic tile combines
 - Analog Compute Engine (ACE™)
 - RISC-V processor, SIMD engine, Scoreboard and SRAM
 - Network-on-Chip (NOC) router
- Each Mythic ACE™
 - Stores weights in non-volatile flash memory
 - Computes matrix multiplication directly inside flash memory array



2-1-8. IEDM 2021

目次

1. 開催概要
2. 2021 IEDM概要
 - 2-1. 参加者の推移
 - 2-2. 論文数の推移
 - 2-3. 地域別論文数の推移
 - 2-4. 機関別論文数
 - 2-5. サブコミッティの構成
3. プログラム
 - 3-1. チュートリアル
 - 3-2. ショートコース
 - 3-3. 基調講演
 - 3-4. フォーカスセッション
 - 3-5. 一般講演
4. 2021 IEDM発表概要
5. 注目論文
 - 5-1. 縦型ナノシートトランジスタ
 - 5-2. 埋込電源配線のメタル材料探索
 - 5-3. 4nm技術
 - 5-4. 2D材料
 - 5-5. 3D積層GaN-Silicon CMOS技術
 - 5-6. 反強誘電性容量を使ったFeRAM

1. 開催概要

1. 開催日時

2021年12月11日 - 15日 (5日間)

日程	内容
12/11	午後にチュートリアル
12/12	終日ショートコース
12/13	本会議開始、午前に基調講演、午後は一般講演
12/14,15	本会議、終日一般講演

2. 開催形式

現地開催*1 + オンデマンドビデオ*2

*1 チュートリアル、ショートコース、基調講演の一部は講演者が現地でライブで講演を行ったが、一般講演はすべて事前収録したビデオが会場で放映された。チュートリアル、ショートコース、基調講演でも講演者が現地に行けない場合は事前収録ビデオが用いられた。

*2 現地開催後の12月17日から現地参加できなかった人のためにオンデマンドビデオが公開された。



URL <https://www.ieee-iedm.org/>

3. 主催

IEEE Electron Devices Society(EDS)

4. IEDM概要

電子デバイス分野の最高峰国際会議

- ・VLSI分野では6月にVLSI Symposiumも有名
- ・ただしIEDMはVLSIデバイス以外の単体デバイスも含む

毎年12月に米国サンフランシスコで開催されている。今年で67回目の開催となる。

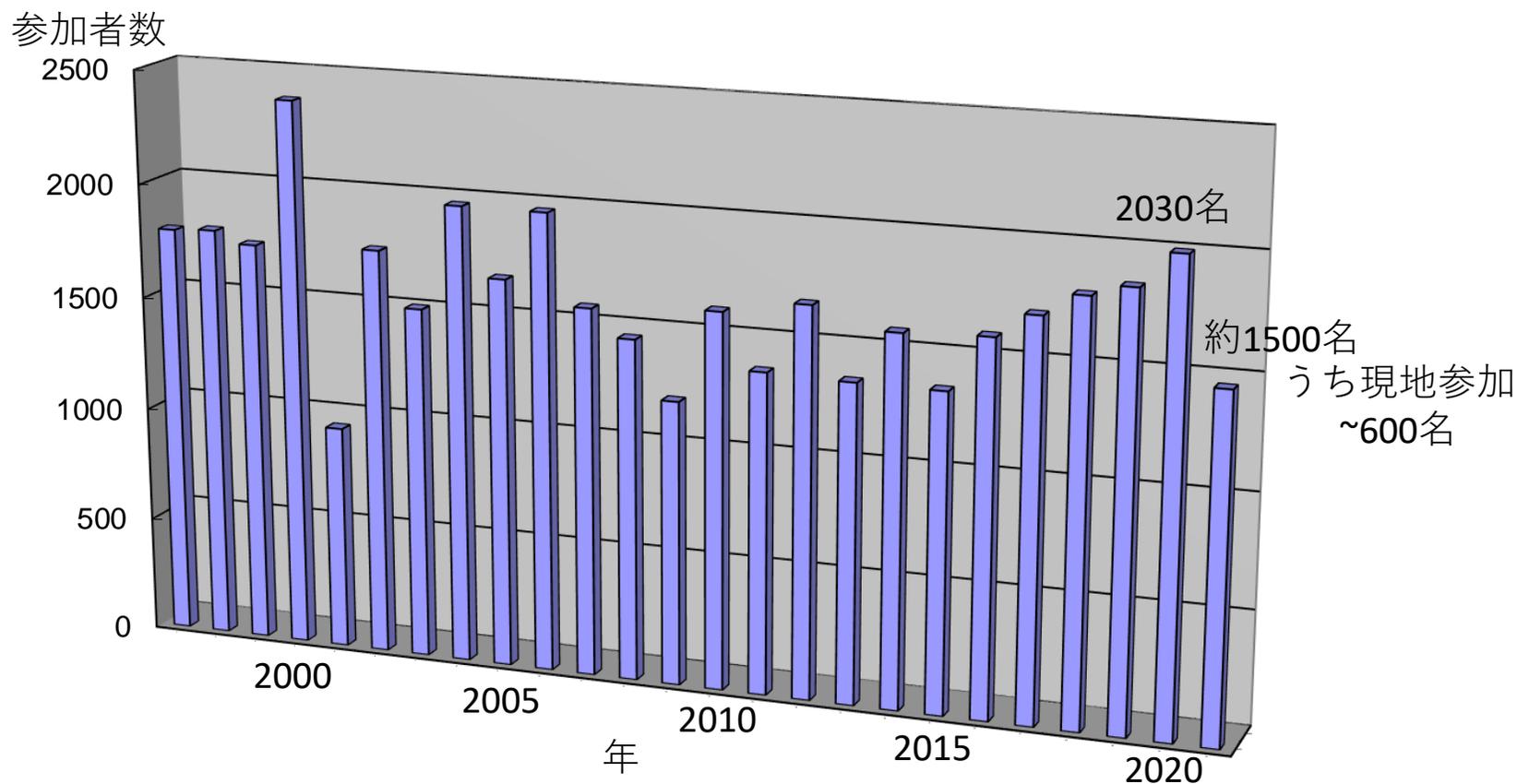
参加者は例年2000名程度。論文数は約230件、採択率~30%。

2. 2021 IEDM概要

2-1. 参加者数の推移

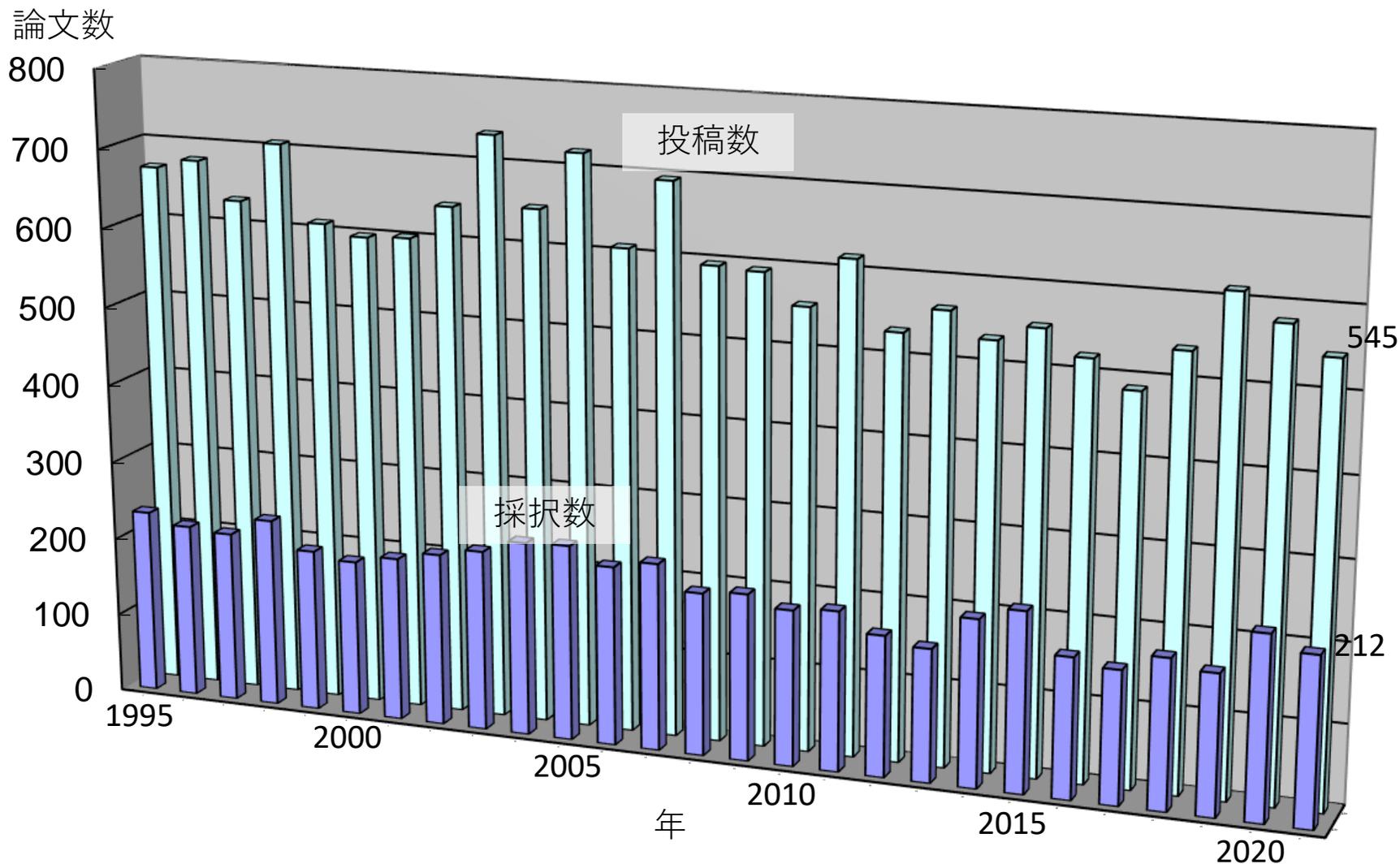
2021 IEDMの参加者数は、会議事務局の非公式発表によれば約1500名であった模様。内訳は現地参加約600名、会議終了後のオンデマンドビデオ視聴による参加者約900名となっている。

プレナリーセッションでの発表に基づいて作成した会議参加者数の推移を以下に示す。2009年のリーマンショック後、徐々に参加者数が回復していた。2020年は完全オンライン開催となり前年に比べて8%程度参加者は増えた。2021年は現地参加者にオンデマンド参加者を加えてもトータルの参加者は減った模様。



2-2. 論文数の推移

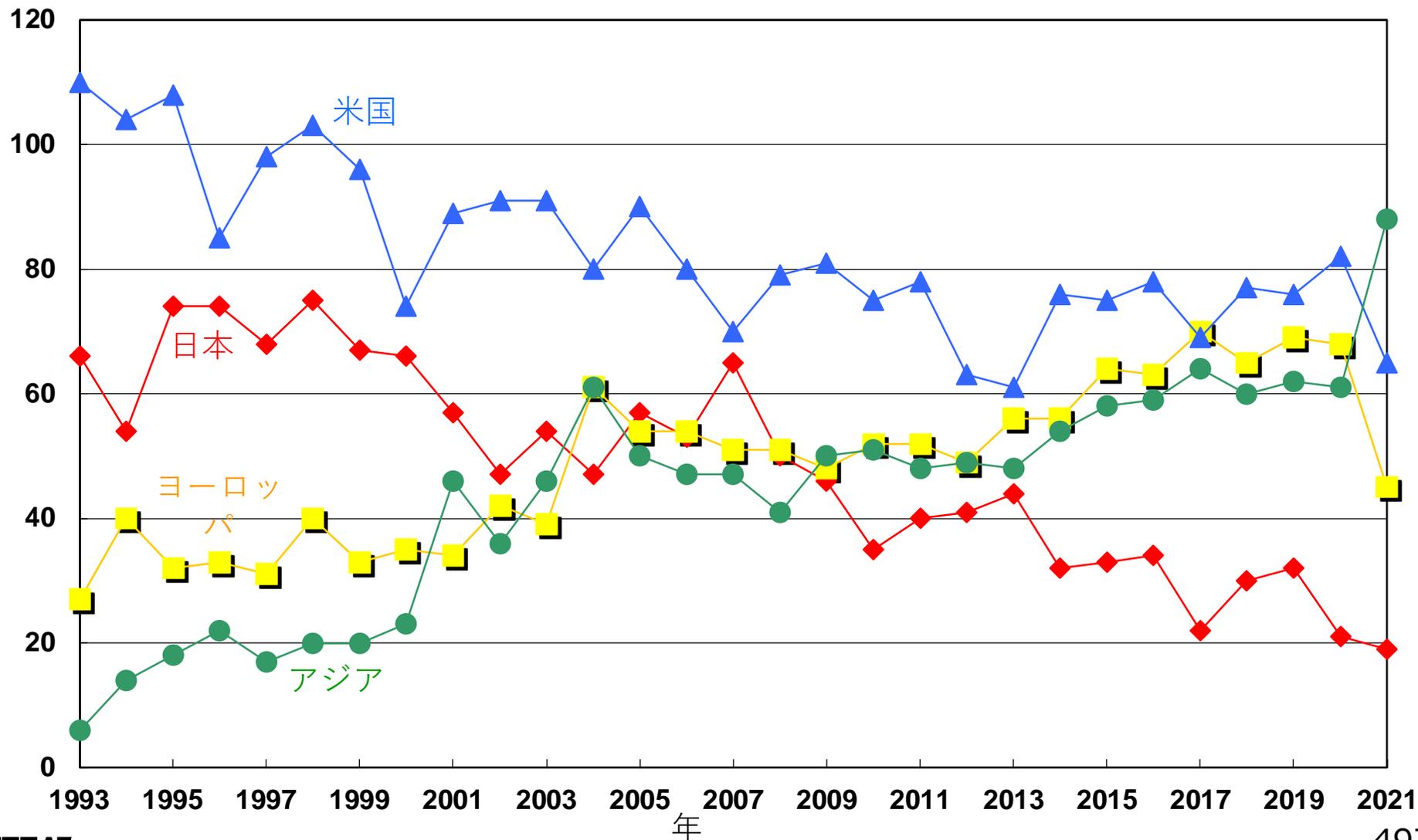
プレナリーセッションでの発表に基づいて作成した論文数の推移を以下に示す。年によって招待講演を含んだ年と含まない年がある。2019-2021年は招待講演を含んでいない。2020年および2021年の投稿数は前年比約10%減となっている。恐らくCOVID-19の影響であると思われる。



2-3. 地域別論文数の推移

発表論文を地域別にカウントして作成した地域別論文数の推移を以下に示す。日本の論文数の減少傾向が明瞭に表れている。2021年はアジア（日本以外）が急激に伸びヨーロッパが減少した。アジアの急伸は中国の論文数が増えたためである。

論文数



2-4. 機関別論文数

発表論文を機関別にカウントして作成した表を以下に示す。日本の機関を青色で示した。今回もimecが断トツの1位となっている。SamsungとIBMも上位の常連。中国科学院と北京大学が論文数を増やしていることがわかる。

2018年		2019年		2020年		2021年	
数	機関名	数	機関名	数	機関名	数	機関名
19	IBM	24	imec	25	imec	19	imec
18	imec	14	Intel	11	Intel	10	Samsung
9	北京大学	14	IBM	11	IBM	9	IBM
8	Samsung	8	CEA LETI	8	Samsung	8	中国科学院
7	ST Microelectronics	7	TSMC	7	ジョージア工科大	7	TSMC
6	東京大学	7	国立交通大学	7	中国科学院	7	北京大学
6	産総研	7	北京大学	6	東京大学	7	KAIST
6	パデュー大学	6	東京大学	6	TSMC	5	Intel
5	Intel	6	キオクシア	6	国立交通大学	5	CEA LETI
5	清華大学	6	Samsung	6	CEA LETI	5	UCバークレー
5	UCサンタバーバラ	6	中国科学院	5	MIT	5	スタンフォード大
4	中国科学院	6	グルノーブル大学	4	北京大学	4	国立陽明交通大学
4	ジョージア工科大	5	ノートルダム大学	4	Macronix	4	ジョージア工科大
4	National Device Lab	4	ソニー	3	ノートルダム大学	4	ノートルダム大学
4	CEA LETI	4	MIT	3	KU Leuven	3	ソニーセミコンダクタ
4	TSMC	4	EPFL	3	Virginia Polytechnic	3	その他9機関
		4	UCサンタバーバラ				

2-5. サブコミッティの構成

IEDMの論文採択はサブコミッティごとに行われる。2019年にサブコミッティが再編され、その後は変わっていない。現在ある9つのサブコミッティを以下に示す。

先端ロジック技術はALT、不揮発性メモリなどはMT、新概念によるデバイス等はEDTで議論される。

ALT Advanced Logic Technology

EDT Emerging Devices and Compute Technology

MT Memory Technology

MS Modeling and Simulation

MAT Microwave, Millimeter Wave, and Analog Technology

ODI Optoelectronics, Displays, and Imaging Systems

PDS Power Devices and Systems

RSD Reliability of Systems and Devices

SMB Sensors, MEMS, and Bioelectronics

3. プログラム

3-1. チュートリアル (1/2)

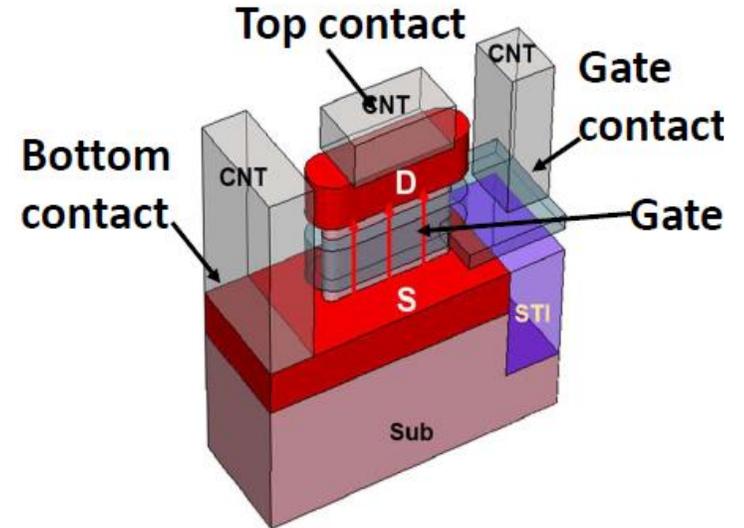
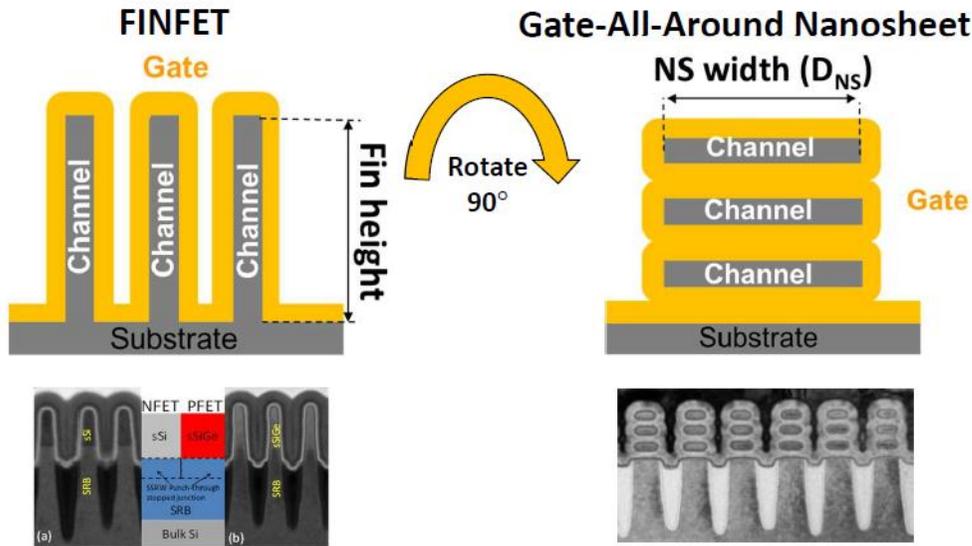
本会議2日前の土曜日の午後に6つのチュートリアル講演が組まれた。一人の講師が最新のトピックスについて基礎から解説する。キーワードを青色で示した。IBMがFinFET以降のデバイス技術について解説した。他に、機械学習やGaNパワーデバイスなどのトピックスもチュートリアル講演で取り上げられた。

#	講演タイトル	講演者/所属
1	Beyond FINFET era: Challenges and opportunities for CMOS technology	Kai Zhao/ IBM
2	TCAD based DTCO and STCO	Asen Asenov/ University of Glasgow
3	6G Technology Challenges from devices to wireless systems	Aarno Pärssinen/ Oulu University
4	Selective and Atomic Scale Processes for Advanced Semiconductor Manufacturing	Robert Clark/ TEL
5	Machine Learning for Semiconductor Device and Circuit Modeling	Elyse Rosenbaum/ Univ. of Illinois at Urbana-Champaign
6	GaN Power Device Technology and Reliability	Dong Seup Lee/ Texas Instruments

3-1. チュートリアル (2/2)

IBMのチュートリアル講演。なぜFinFETからナノシートにデバイスアーキテクチャが変わるのか、どのようなプロセス・デバイス上の課題があるかをわかりやすく解説した。後述の縦型ナノシートトランジスタにも言及している。

1. Beyond FinFET Era (IBM)



VTFET (Vertical Transport NanoSheet FET)

(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

3-2. ショートコース (1/4)

本会議前日の日曜日に2つのショートコースが組まれた。それぞれ6名の講師からなる。例年、テーマはロジックとメモリに分かれるが、今回はロジックのほかにエッジ向け技術が取り上げられた。キーワードを青色で示した。

Short Course 1: Future scaling and integration technology

#	講演タイトル	講演者/所属
1	Processes and Materials Engineering Innovations for Advanced Logic Transistor Scaling	Benjamin Colombeau/ Applied Materials
2	Interconnect Resistivity: New Materials	Daniel Gall/ Rensselaer Polytechnic Institute
3	Metrology and Material Characterization for the Era of 3D Logic and Memory	Roy Koret/ Nova Ltd.
4	Beyond FinFET Devices: GAA, CFET, 2D Material FET	Chung-Hsun Lin/ Intel
5	Heterogeneous Integration Using Chiplets & Advanced Packaging	Madhavan Swaminathan/ Georgia Tech
6	Design-Technology Co-Optimization / System-Technology Co-Optimization	Victor Moroz/ Synopsys

3-2. ショートコース (2/4)

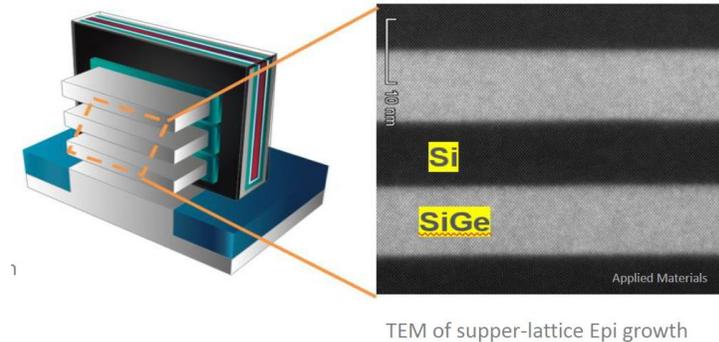
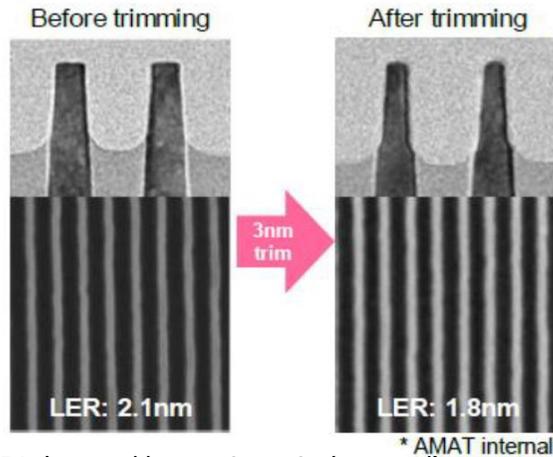
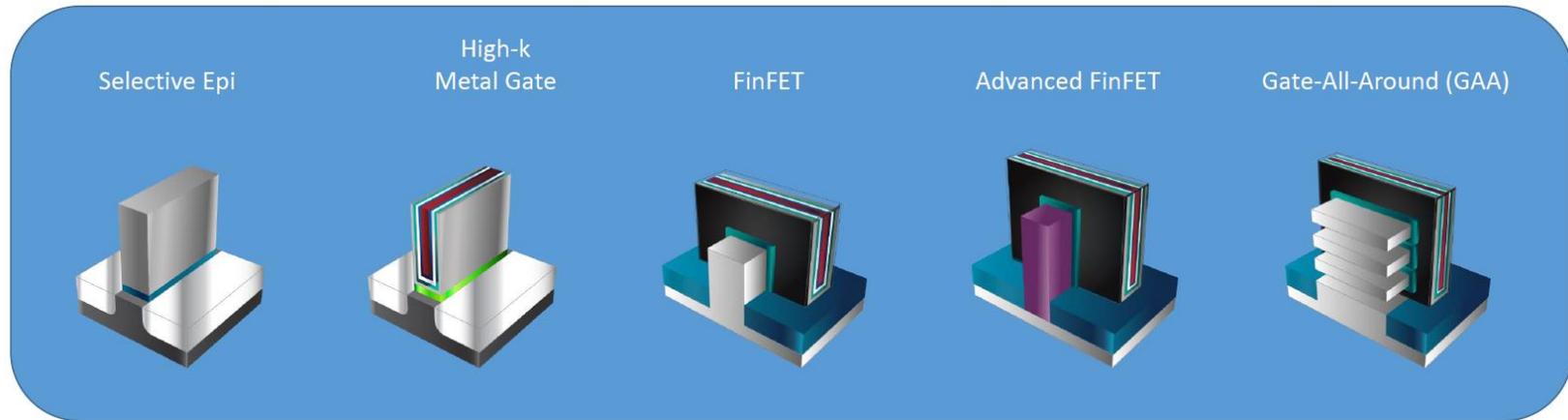
Short Course 2: Emerging Technologies for Low Power Edge Computing

#	講演タイトル	講演者/所属
1	Mobile NPU for Intelligent Human Computer Interaction	Hoi-Jun Yoo/ KAIST
2	Brain-inspired strategies for optimizing the design of neuromorphic sensory-processing systems	Giacomo Indiveri/ University of Zurich
3	Memory based AI & Data Analytics solutions	Euicheol Lim/ SK Hynix
4	Material strategies for memristor-based AI hardware and their heterointegration	Jeewan Kim/ MIT
5	RRAM devices for data storage and in-memory computing	Wei Lu/ University of Michigan
6	Practical implementation of wireless power transfer	Hubregt Visser/ IMEC

3-2. ショートコース (3/4)

ショートコース1の最初の講演は製造装置メーカー最大手のApplied Materialsの講演であった。装置メーカーの実力がよくわかる内容であった。Finの形状制御、界面準位低減、S/D形成、EOTスケーリング、マルチVth制御、コンタクト抵抗低減、ばらつき低減など、さまざまなプロセス課題と対応策を解説した。

Short Course 1: 1. Advanced Logic Transistor Scaling (Applied Materials)

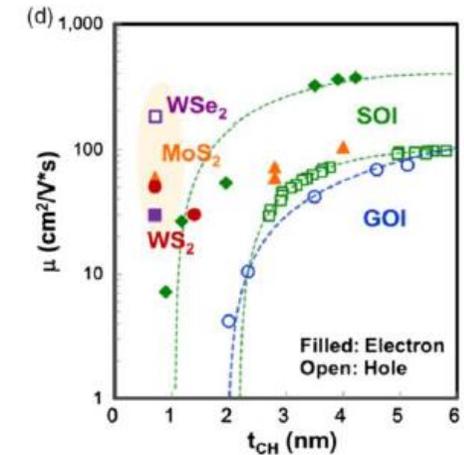
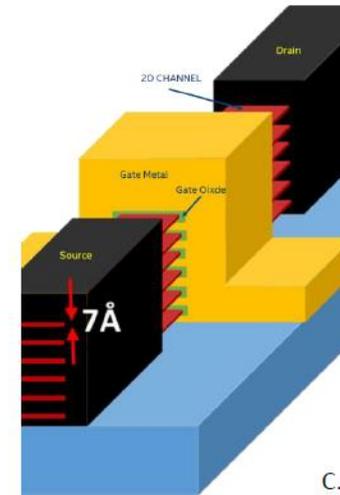
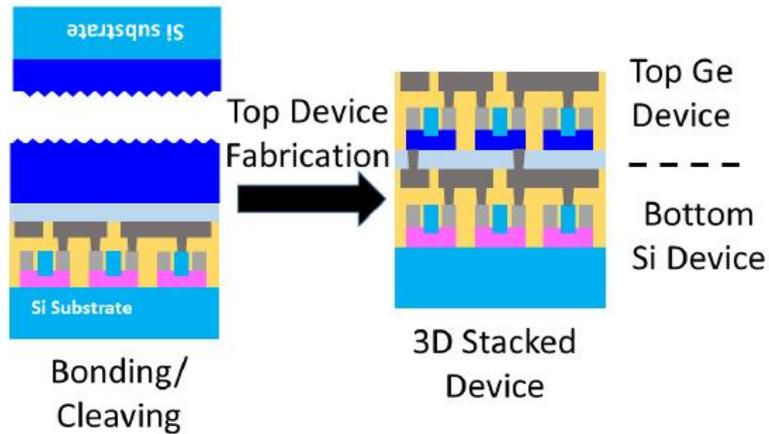


(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

3-2. ショートコース (4/4)

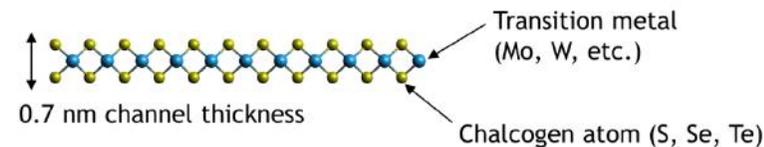
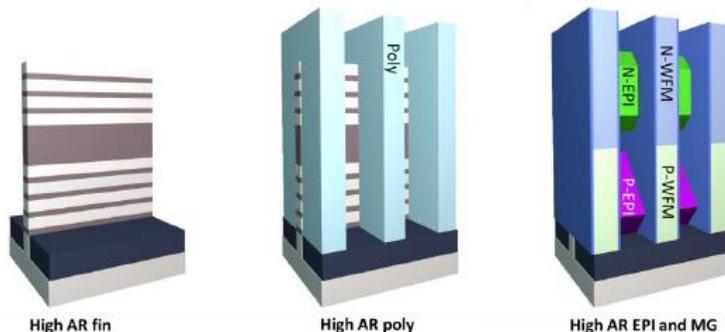
インテルのショートコース講演。インテルとしてはムーアの法則維持のため、デバイススケールをさらに進めたい。FinFETの次としてナノシート構造および3次元構造デバイスを挙げている。3次元ではシーケンシャルとモノリシックの2案がある。前者はSi上のGaNデバイスの集積などに有利であり、後者はCFET等のロジック・メモリ応用に有利。さらに2次元材料についても精力的に研究を行っているが、まだ候補となる材料を絞り切れていない模様。

Short Course 1: 4. Beyond FinFET Devices: GAA, CFET, 2D Material FET (Intel) Sequential 3D



C. Dorow et al, VLSI 2021

Self-Aligned (Monolithic) 3D



2D-Material Channel FET for Ultimate Scaling

(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

3-3. 基調講演

- 基調講演は例年通り3名であった。
- SamsungのKinam Kimは2010年に続いて2回目の基調講演となった。前回は、DRAM、NANDフラッシュ等のメモリが中心の講演だったが、今回はロジックやイメージセンサも含めて半導体による明るい未来を描いた講演であった。（事前収録によるビデオ講演）
- 2人目はMeta社（旧Facebook社）の拡張現実に関する講演。VRやARを含むXR技術が今後50年にわたってコンピューティングの主流技術になると予想した。（事前収録によるビデオ講演）
- 3人目はIBMによる量子コンピューティングの講演であった。量子コンピューティングは前回は基調講演のテーマであったので、IEDMが量子コンピューティングを重要視していることがよくわかる。（現地におけるライブ講演）

#	講演タイトル	講演者/所属
1	The Smallest Engine Transforming Humanity: The Past, Present, and Future	Kinam Kim/ Samsung Electronics
2	Creating the Future: Augmented Reality, the next Human-Machine Interface	Michael Abrash/ Reality Labs, Meta
3	Quantum Computing Technology	Heike Riel/ IBM Quantum

3-4. フォーカスセッション（1/3）

- 今回は5つのフォーカスセッションが組まれた（前回は6セッション）。フォーカスセッションは、最新の重要トピックスを事前に設定してセッションを構成するもので、**講演はすべて招待講演**である。
- 今回は、複数のサブコミッティにまたがるフォーカスセッションが多かった（前回はすべて単独のサブコミッティであった）。
- ロジック関連(ALT)では、三次元集積化が最新の技術トレンドであり、モノリシック集積化から実測技術まで、さまざまな3D技術が発表された。**次スライドで詳しく紹介**する。
- EDTの量子コンピューティングは、2018年以降連続で取り上げられている（前回はCryogenic Electronics）。
- EDT/ODIのトポロジカル材料がフォーカスセッションで取り上げられるのは恐らく初めて。**新材料・新物性への期待**が込められている。

#	サブコミッティ	セッションタイトル
1	EDT	Device Technology for Quantum Computing
2	MT/MS/ALT	STCO for Memory-Centric Computing and 3D Integration
3	ALT	Stacking of Devices, Circuits, Chips: Design, Fabrication, Metrology – Challenges and Opportunities
4	SMB/ODI	Technologies for VR and Intelligence Sensors
5	EDT/ODI	Topological Materials, Devices, and Systems

※キーワードを青色で示した

3-4. フォーカスセッション (2/3)

- 3D積層・3D集積化のフォーカスセッション。
- 東京エレクトロン(TEL)がCFETデバイス設計の講演を行ったことは注目される。製造装置メーカーの技術力が示されている。
- 一方、日本の日立ハイテクが複雑化する3D構造デバイスの評価に関する講演を行った。
- 3D集積技術の多くは3D実装技術である。多くの3D実装技術が発表された。

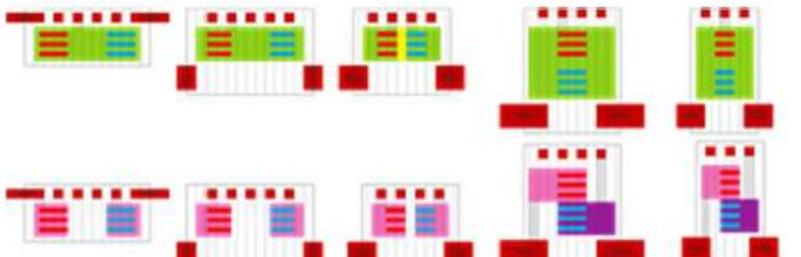
Session 3: ALT: Stacking of Devices, Circuits, Chips: Design, Fabrication, Metrology Focus Session

#	講演タイトル	講演機関
3.1	CFET Design Options, Challenges, and Opportunities for 3D Integration	TEL
3.2	3D Sequential Integration: Applications and Associated Key Enabling Modules (Design & Technology)	CEA-LETI
3.3	Inspection and Metrology Challenges for 3nm Node Devices and Beyond	Hitachi High-Tech
3.4	Heterogeneous Integration Enabled by the State-of-the-Art 3DIC and CMOS Technologies: Design, Cost, and Modeling	Synopsys
3.5	Design for 3D Stacked Circuits	North Carolina State University
3.6	3D SoC Integration, Beyond 2.5D Chiplets	imec
3.7	Foundry Perspectives on 2.5D/3D Integration and Roadmap	TSMC

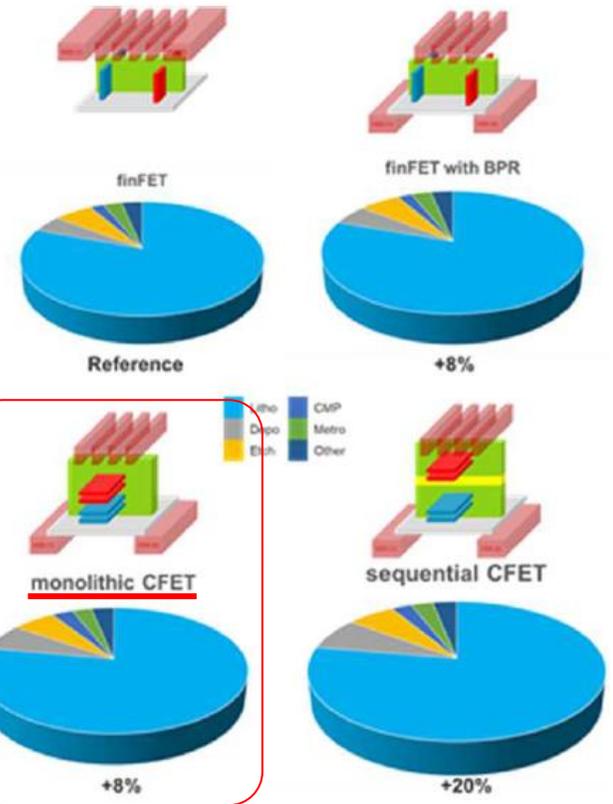
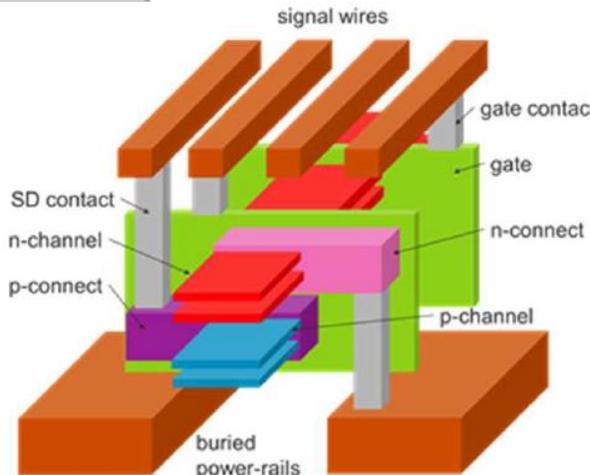
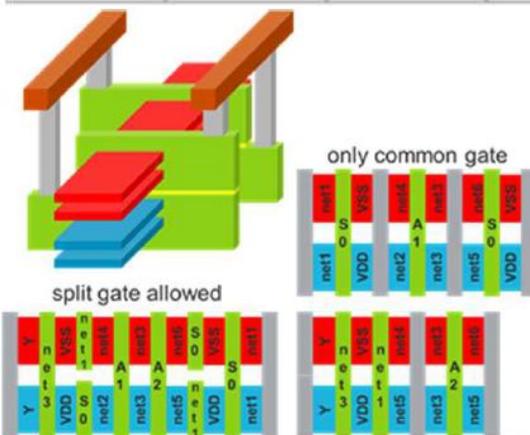
※キーワードを青色で示した

3-4. フォーカスセッション (3/3)

- 東京エレクトロン（米国オルバニー）からのCFETスタンダードセル設計に関する招待講演。製造装置メーカーがスタンダードセル設計まで検討していることがわかる。
- FinFET, NSからCFETへと発展することでスタンダードセルの高さを6.5Tから3.5Tにできる。CFETのプロセスコスト増はBPRと同程度と試算。SequentialプロセスはCFETではコスト上不利であり、Monolithic CFETの開発が必須と結論している。



6.5T NS	6.5T NS BPR	5T FS BPR	5T CFET BPR	3.5T CFET
FET W = 33nm	FET W = 27.5nm	FET W = 22nm	FET W = 33nm	FET W = 22nm
n-to-p S = 44nm	n-to-p S = 44nm	n-to-p S = 11nm	n.a.	n.a.
Power W = 44nm	Power W = 22nm	Power W = 33nm	Power W = 55nm	Power W = 33nm



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

3-5. 一般講演 (1/3)

- 一般講演はロジック技術を中心にセッションを紹介する。
- Session 26は、3nm世代の本命構造と言われているナノシートの論文を集めたセッション。IBMやimecが精力的にナノシートの開発を進めている。
- imecは、2021年6月のVLSIシンポジウムでナノシートを進化させたフォークシート構造のトランジスタを発表済み。nFETとpFETのスペースを縮められる。26.2ではフォークシートとナノシートのDC性能を比較し、フォークシートのDC性能はナノシートと比較して遜色がないことを示した。

Session 26: ALT: Advances in Nanosheet FET Devices

※#が赤字の講演は注目論文として後述する

#	講演タイトル	講演機関
26.1	Vertical-Transport Nanosheet Technology for CMOS Scaling beyond Lateral-Transport Devices	IBM
26.2	Comparison of Electrical Performance of Co-Integrated Forksheets and Nanosheets Transistors for the 2nm Technological Node and Beyond	imec
26.3	Critical Elements for Next Generation High Performance Computing Nanosheet Technology	IBM
26.4	Highly Stacked 8 Ge _{0.9} Sn _{0.1} Nanosheet pFETs with Ultrathin Bodies (~3nm) and Thick Bodies (~30nm) Featuring the Respective Record I _{ON} /I _{OFF} of 1.4x10 ⁷ and Record I _{ON} of 92μA at V _{OV} =V _{DS} = -0.5V by CVD Epitaxy and Dry Etching	National Taiwan University
26.5	Gate-Last I/O Transistors based on Stacked Gate-All-Around Nanosheet Architecture for Advanced Logic Technologies	IBM

※キーワードを青色で示した

3-5. 一般講演 (2/3)

- 従来、配線の論文はIEDMでは比較的少なかったが、特にロジック技術における配線の重要性が増し、2020年にはALTで配線のフォーカスセッションが組まれた。2021年は配線の一般セッションが組まれるまでに配線の論文数が増加した。
- 注目は埋込電源配線(BPR)と裏面配線である。これらを採用するとスタンダードセルの大幅面積縮小が可能となる。22.4でimecが裏面配線を用いたSRAMマクロとロジック回路について講演した。22.5のimecのBPRメタル探索は注目論文。
- 5nm未満の技術ノードでは、製造コスト増大のため3D的アプローチが必須となる。22.3ではQualcommがチップをチップレットに分けて3D集積する“partition”アプローチと、追加機能を3D集積する“Add-on”アプローチの長短を比較している。

Session 22: ALT: Front and Back-side Interconnects

※#が赤字の講演は注目論文として後述する

#	講演タイトル	講演機関
22.1	Fully Self-Aligned via Integration for Interconnect Scaling Beyond 3nm Node	TSMC
22.2	Electromigration and Line R of Graphene Capped Cu Dual Damascene Interconnect	IBM
22.3	System Design Technology Co-Optimization for 3D Integration at <5nm nodes	Qualcomm
22.4	Design and Optimization of SRAM Macro and Logic Using Backside Interconnects at 2nm node	imec
22.5	Buried Power Rail Metal Exploration Towards the 1nm Node	imec

※キーワードを青色で示した

3-5. 一般講演 (3/3)

- Session 13はロジックで最も注目されるCMOSプラットフォームのセッション。今年はゲートスタックと組み合わせたセッションとなった。
- 注目された3nm技術の発表はなかった。
- 13.2のCold CMOSは77K動作のCMOSの発表であった。Cold CMOSは性能向上が期待され、近年再評価されつつある。
- 13.3の4nm技術は、現在までに学会で報告された中では最も微細な技術である。

Session 13: Platform Technologies and Gate Stack

※#が赤字の講演は注目論文として後述する

#	講演タイトル	講演機関
13.1	18nm FDSOI Technology Platform Embedding PCM & Innovative Continuous-Active Construct Enhancing Performance for Leading-Edge MCU Applications	Samsung
13.2	Design Technology Co-Optimization for Cold CMOS Benefits in Advanced Technologies	TSMC
13.3	High Performance 4nm FinFET Platform (4LPE) with Novel Advanced Transistor Level DTCO for Dual-CPP/HPHD Standard Cells	Samsung
13.4	Al-doped and Deposition Temperature-engineered HfO ₂ Near Morphotropic Phase Boundary with Record Dielectric Permittivity (~68)	National University of Singapore
13.5	Dipole-First Gate Stack as a Scalable and Thermal Budget Flexible Multi-Vt Solution for Nanosheet/CFET Devices	imec
13.6	Demonstration of Low EOT Gate Stack and Record Transconductance on Lg=90 nm nFETs Using 1.8 nm Ferroic HfO ₂ -ZrO ₂ Superlattice	UC Berkeley

※キーワードを青色で示した

4. 2021 IEDM発表概要

● 電子デバイス領域での世界の流れ

- ・ 先端CMOSロジックでは、5nmまでFinFET構造が採用されてきた。
- ・ 2022年中量産とされる3nmでどのような構造が採用されるか注目される。
- ・ チップレットなど実装技術による3D集積化が進みつつある。

● IEDMでの技術動向

- ・ 次世代デバイス候補として**ナノシート構造・CFET**等の発表が増えている。
- ・ **埋込配線等の新規配線技術**もロジックの重要技術として発表が増えている。
- ・ **HfO₂系の強誘電体によるトランジスタ・メモリ**の発表は今回も非常に多い。

● 将来的な研究開発の方向性

- ・ 企業も1原子層の厚さを有する**2次元材料トランジスタ**に注目している。
- ・ **In-Memory Computing**など**新しいコンピューティング法**が模索されている。
- ・ さらに**量子コンピューティング**への期待も大きい。

5. 注目論文

5-1. 縦型ナノシートトランジスタ (1/2)

26.1. Vertical-Transport Nanosheet (IBM)

(1) 本発表が応用される分野、製品

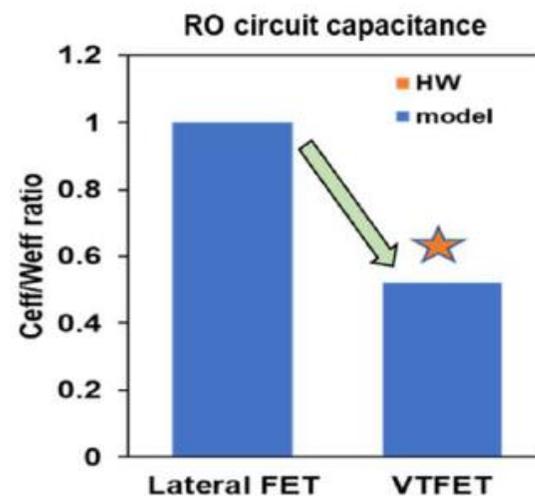
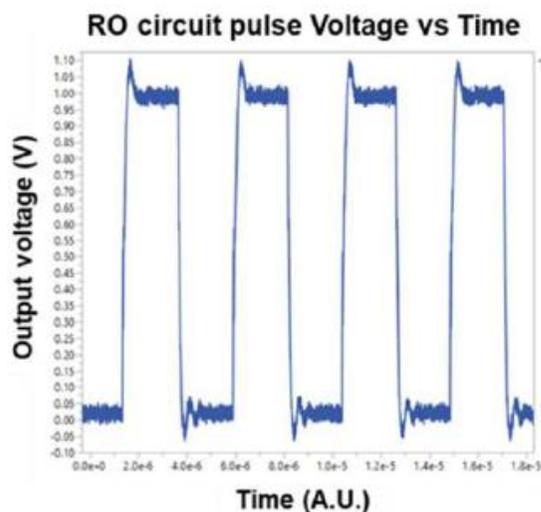
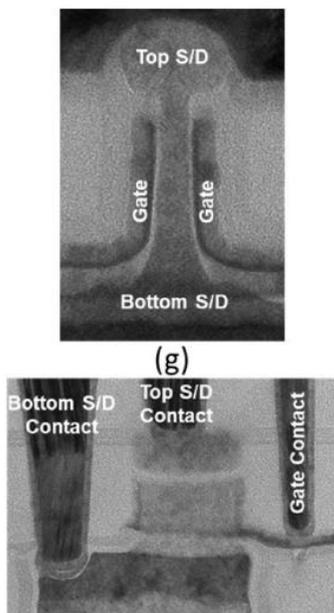
主にCPUなどのロジック製品

(2) 当該分野における技術ニーズ

高集積密度、高速、低消費電力、低コスト

(3) 本発表の概要

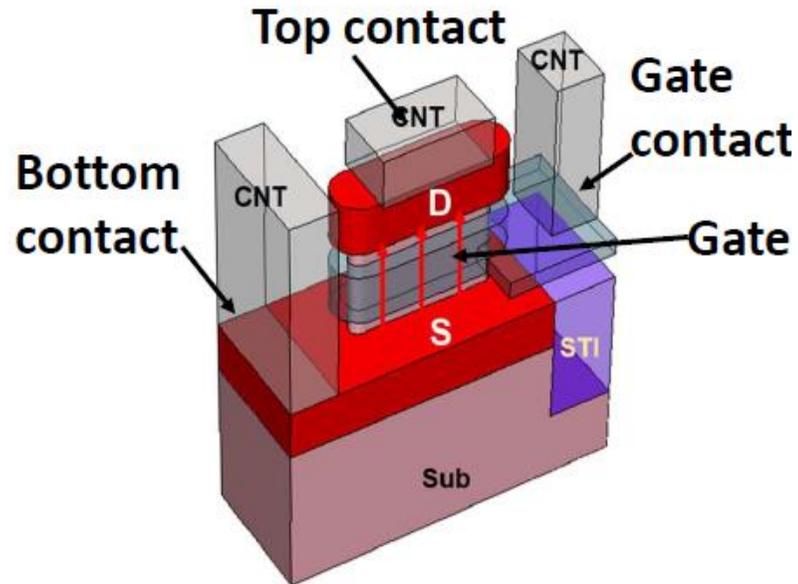
- 縦型FETはimec等で検討されてきたが、本論文はリングオシレータの実測データを示すなど完成度が極めて高い。
- 縦型構造では素子分離のためのスペースが小さくでき、寄生容量も低減されるため集積度・速度でもメリットがあると結論している。
- 従来ロードマップにおけるナノシートからCFET*への流れの延長線上にない技術といえる。今後は3D集積化が可能かどうか焦点となろう。



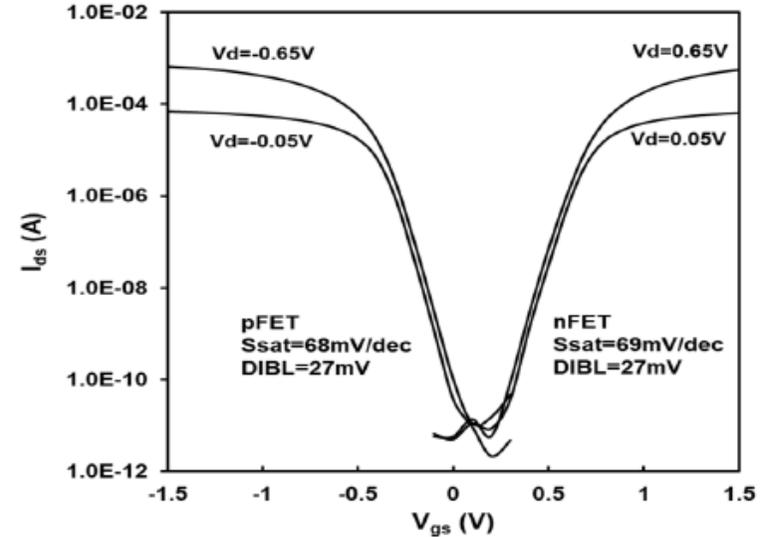
(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

* CFET(Complementary FET) : pチャンネルFETとnチャンネルFETをシリコン表面と垂直に積層したものだ。

5-1. 縦型ナノシートトランジスタ (2/2)



(注)上図は論文には記載されていない。チュートリアル講演の図面から転用した。



(4) 技術的に特筆すべき点

これまで実用化されてこなかった縦型構造トランジスタで従来の横型トランジスタを凌ぐ性能・集積密度を達成している点。しかもリングオシレータ性能を示すなど完成度が高い。

(5) 本発表が注目される理由

IBMからは初めての縦型トランジスタの発表であり、本気で縦型構造に取り組んでいることが分かる点。

(6) 実用化までの課題

集積回路にしたときの性能、プロセスコスト、歩留など

(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-2. 埋込電源配線のメタル材料探索 (1/2)

22.5. Buried Power Rail Metal for 1nm Node (imec)

(1) 本発表が応用される分野、製品

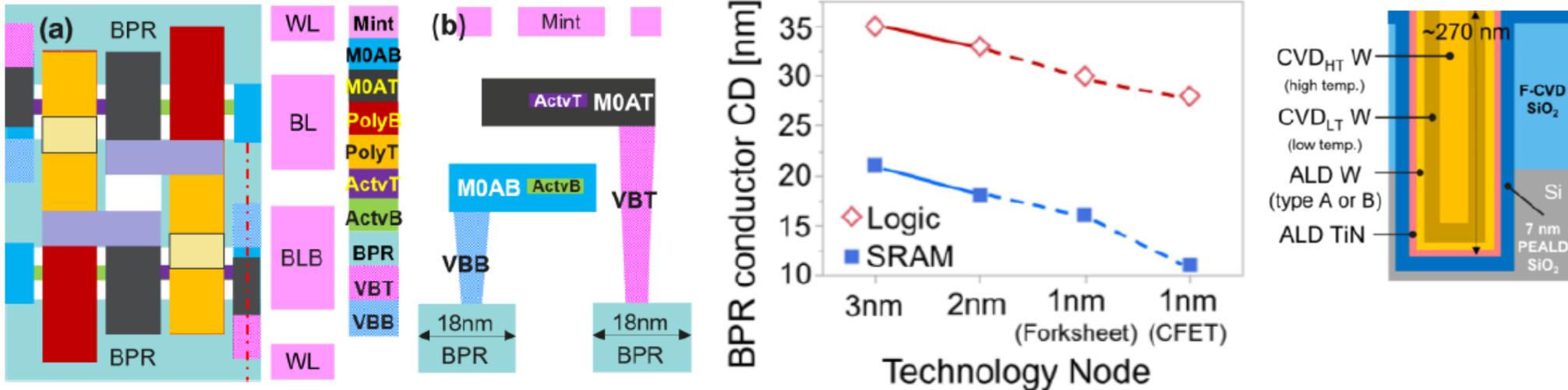
主にCPUなどのロジック製品

(2) 当該分野における技術ニーズ

高集積密度、高速、低消費電力、低コスト

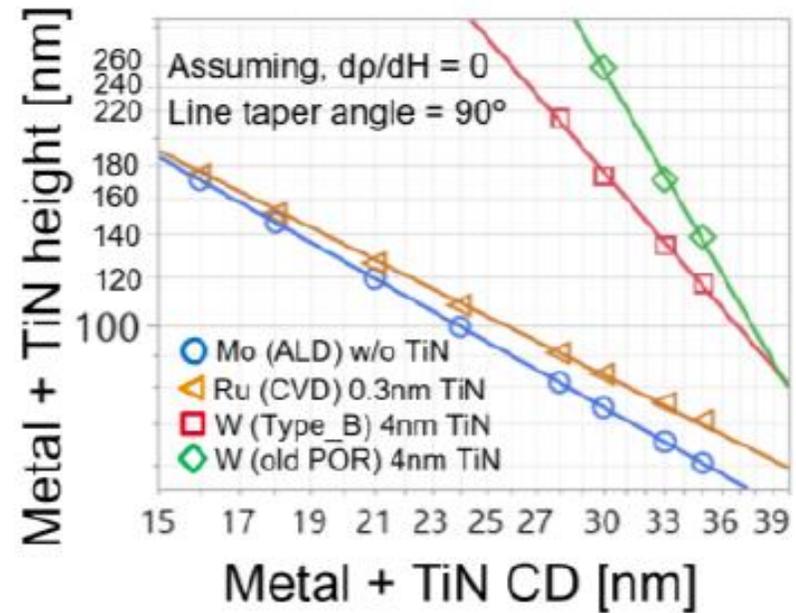
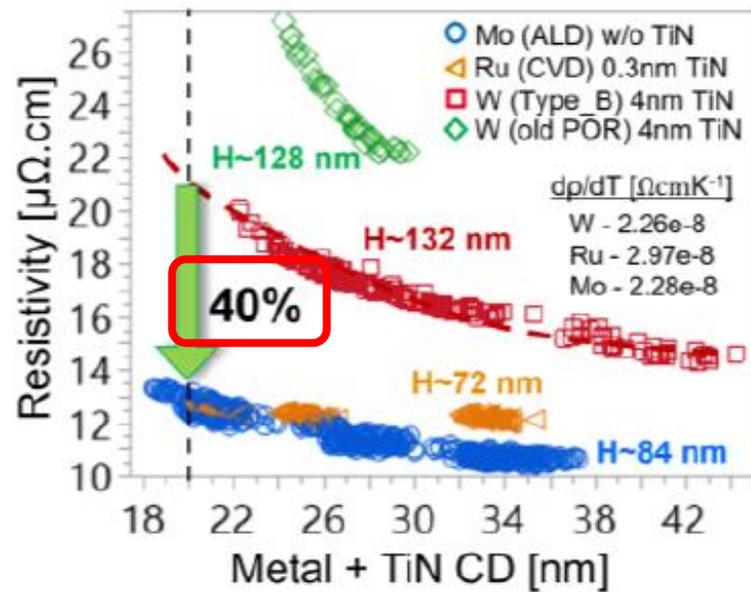
(3) 本発表の概要

- ロジックにおける配線技術の重要性が増している。スタンダードセル面積を縮小するためには埋込電源配線(BPR)は必須である。
- 本論文では、埋込配線材料としてW(タングステン)は抵抗が大きすぎるためRu(ルテニウム)またはMo(モリブデン)が必要と結論している。



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-2. 埋込電源配線のメタル材料探索 (2/2)



(4) 技術的に特筆すべき点

BPR (埋込電源配線) はスタンダードセル面積を小さくし集積密度を上げることを可能とする技術として注目されている。本論文ではその配線材料に注目して、候補となる金属材料を電気特性から明らかにした点。

(5) 本発表が注目される理由

電気特性が示されており、開発がある程度まで進んでいることを示しているため

(6) 実用化までの課題

集積回路にしたときの性能、プロセスコスト、歩留など

(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-3. 4nm技術 (1/3)

13.3 4nm FinFET (Samsung)

(1) 本発表が応用される分野、製品

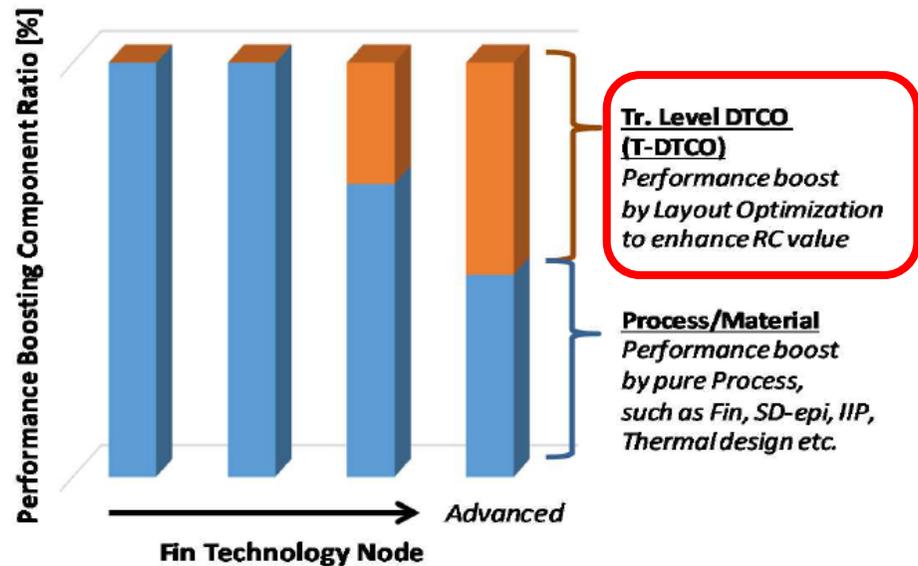
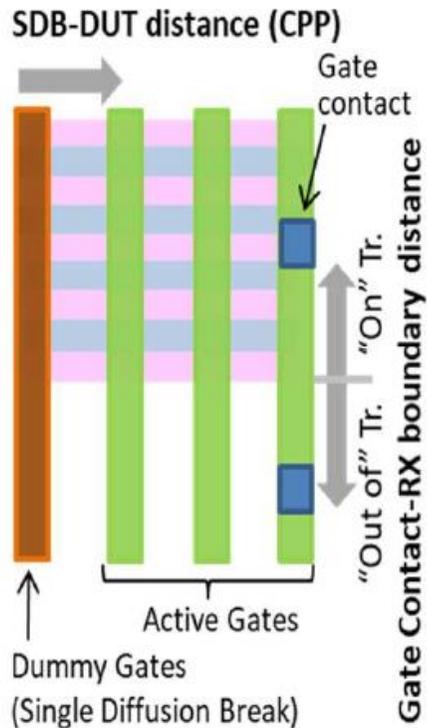
主にCPUなどのロジック製品

(2) 当該分野における技術ニーズ

高集積密度、高速、低消費電力、低コスト

(3) 本発表の概要

- ・ 5nmとの違いはContact-over-RX (アクティブ領域にゲートコンタクトを設けること) が採用されたこと。それ以外には新しい内容は無い。単なるデバイス微細化よりレイアウト最適化による性能向上が主流になっていることが示されている。
- ・ 一部デバイス特性図が示されているが、軸がa.u.(arbitrary Unit: 任意単位)となっており、特性の絶対値は示されていない。

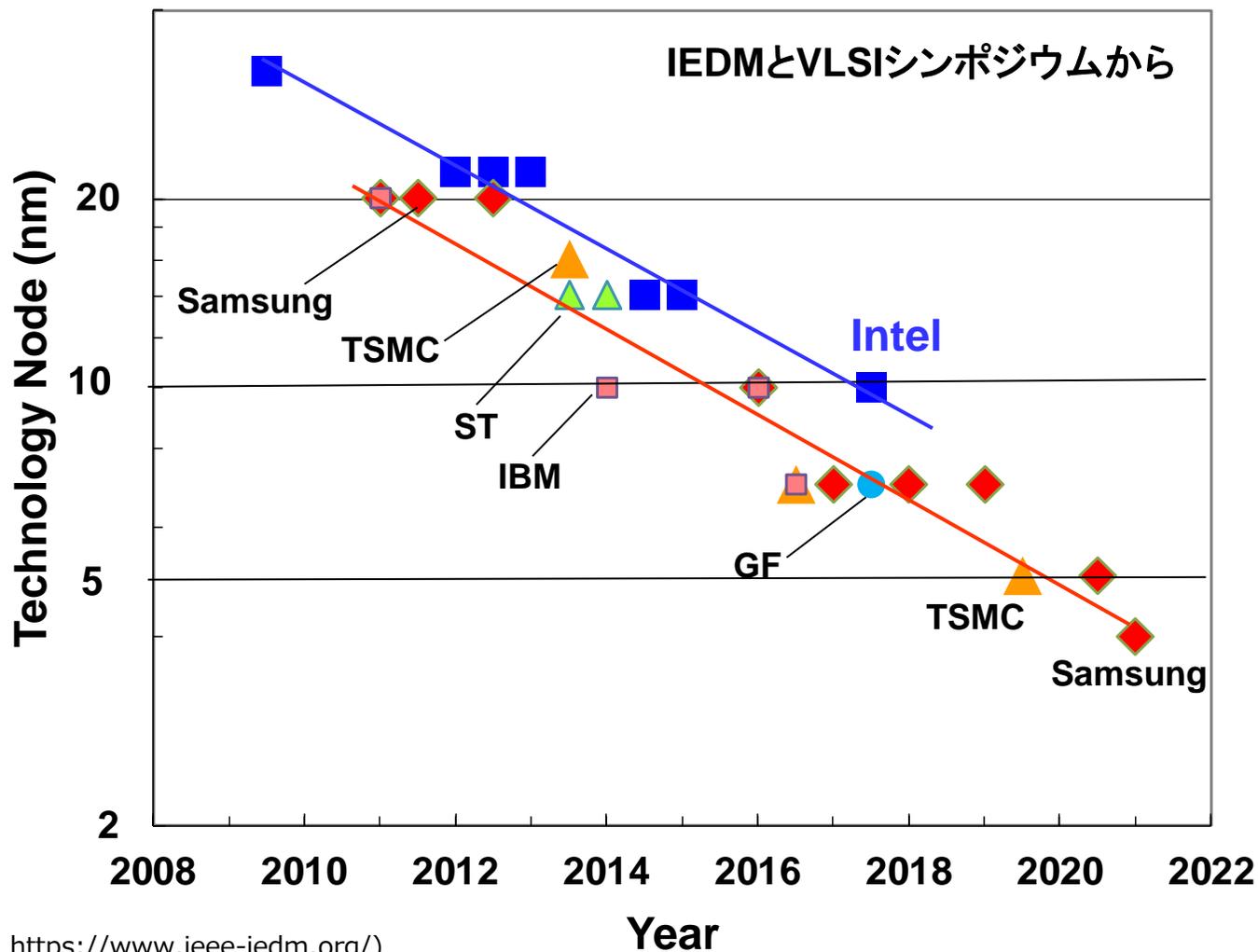


(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-3. 4nm技術 (2/3)

主要国際会議 (IEDMとVLSIシンポジウム) で発表されたCMOSプラットフォーム論文。

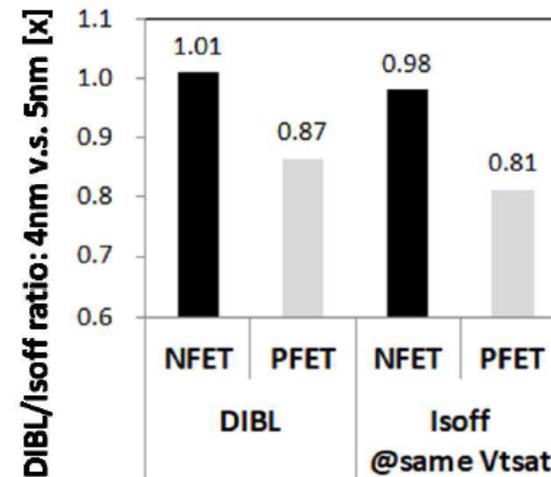
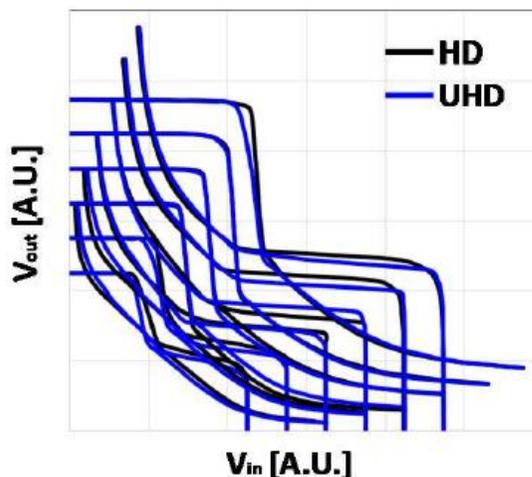
今回初めて4nmがSamsungから発表された。各社は3nm技術の開発を進めているが、少なくとも学会では3nmは現在まで発表されていない。



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-3. 4nm技術 (3/3)

Key Feature	8nm	7nm (2 nd gen.)	5nm (5LPE)	4nm (4LPE)
Fin	4 th gen.	5 th gen.	6 th gen.	7 th gen.
S/D	4 th gen.	5 th gen.	6 th gen.	7 th gen.
Contact /BEOL	MPT	EUV	EUV	EUV
Process	MDB/SDB	MDB/SDB	MDB/SDB	SDB
Future	Dual CPP	Dual CPP	Dual CPP	Dual CPP
			Contact -on-RX edge	Contact -over-RX
Reference	VLSI 2018'	VLSI 2018'	IEDM 2020'	This work



(4) 技術的に特筆すべき点

これまで学会で報告された中で最も技術ノードの数字が小さい技術は5nmだったが、今回初めて学会で4nmが発表された点

(5) 本発表が注目される理由

実際に製品に適応される技術が示されていること

(6) 実用化までの課題

プロセスコスト、歩留など

5-4. 2D材料 (1/2)

7.1. Advancing 2D Monolayer CMOS (Intel)

(1) 本発表が応用される分野、製品

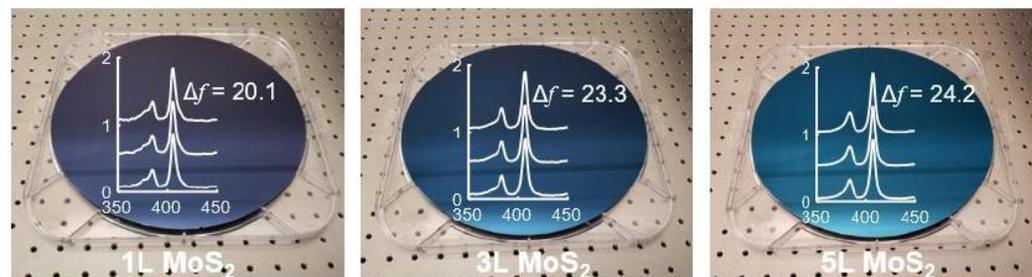
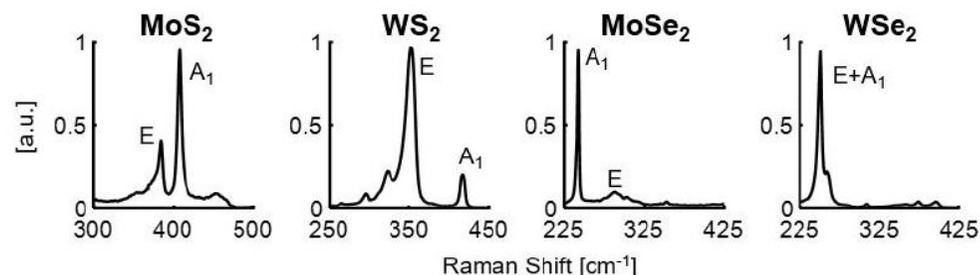
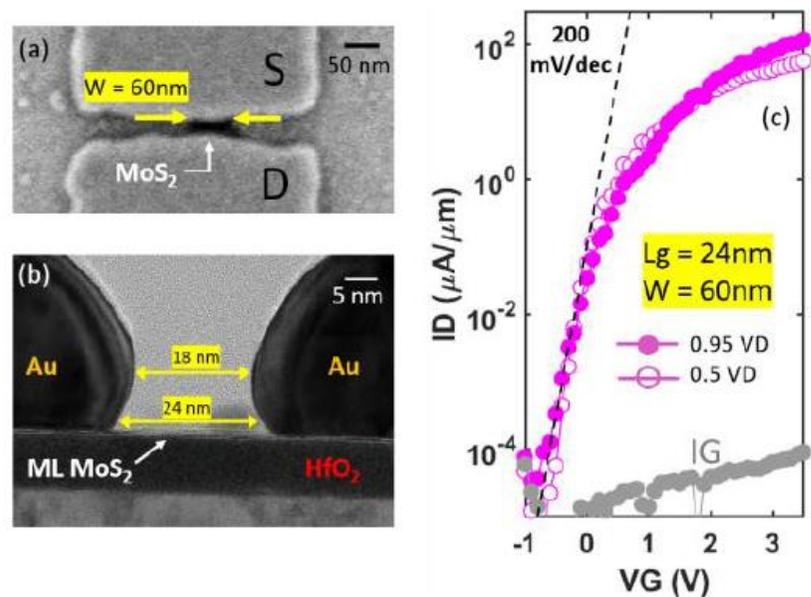
主にCPUなどのロジック製品

(2) 当該分野における技術ニーズ

高集積密度、高速、低消費電力、低コスト

(3) 本発表の概要

- まだ集積回路とは言い難いが、**2D材料(単分子層半導体)**は**将来のデバイスチャネル材料として非常に注目されている**。
- インテルは300mmウェーハ上にCVDでMoS₂等の2D材料を堆積させBEOL相当のプロセス温度でFETを試作した。
- ただし構造はバックゲート構造である。コンタクト抵抗の低減によりpFETとして最高のオン電流を達成した。



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-4. 2D材料 (2/2)

Paper	Intel [4]	[7]	[9]	[10]	[11]	This Work	[12]	[13]
NMOS/PMOS	NMOS					PMOS		
Material	MoS2	MoS2	WS2	MoS2	MoS2	WSe2	WSe2	MoS2
# ML	1	1	~3	1	1	1	1	~7
Dep Method	MBE	SS CVD	MOCVD	SS CVD	SS CVD	SS CVD	SS CVD	CVD
Single or Dual Gate	single	single	dual	dual	dual	single	single	single - Ushaped
I _{max} (@ 1V _{ds}) [μA/μm]	330	850	310	260	240 (@1.6V _{ds})	50	2	120
SS [mV/decade]	89	880	117	250	73	141	97	260

(4)技術的に特筆すべき点

量産性に優れたCVDで300mmウェーハに2D材料を成膜してデバイス特性を評価している点。

(5)本発表が注目される理由

インテル社が本気で2D材料の開発に取り組んでいることがわかるため

(6)実用化までの課題

まずは、Siに匹敵する性能（ドレイン電流、サブスレッショルド係数など）を出すこと。実用化はまだ遠い。

5-5. 3D積層GaN-Silicon CMOS技術 (1/2)

11.1. 3D Layer Transfer GaN-Silicon Finfet CMOS Integration (Intel)

(1) 本発表が応用される分野、製品

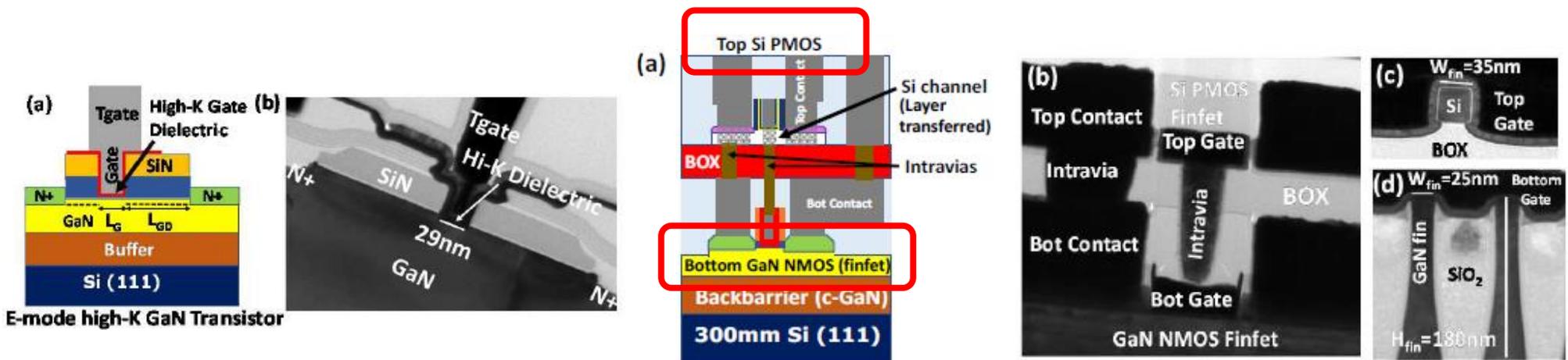
パワーエレクトロニクス製品、パワーIC

(2) 当該分野における技術ニーズ

高耐圧、低オン抵抗、高信頼性、集積回路との集積

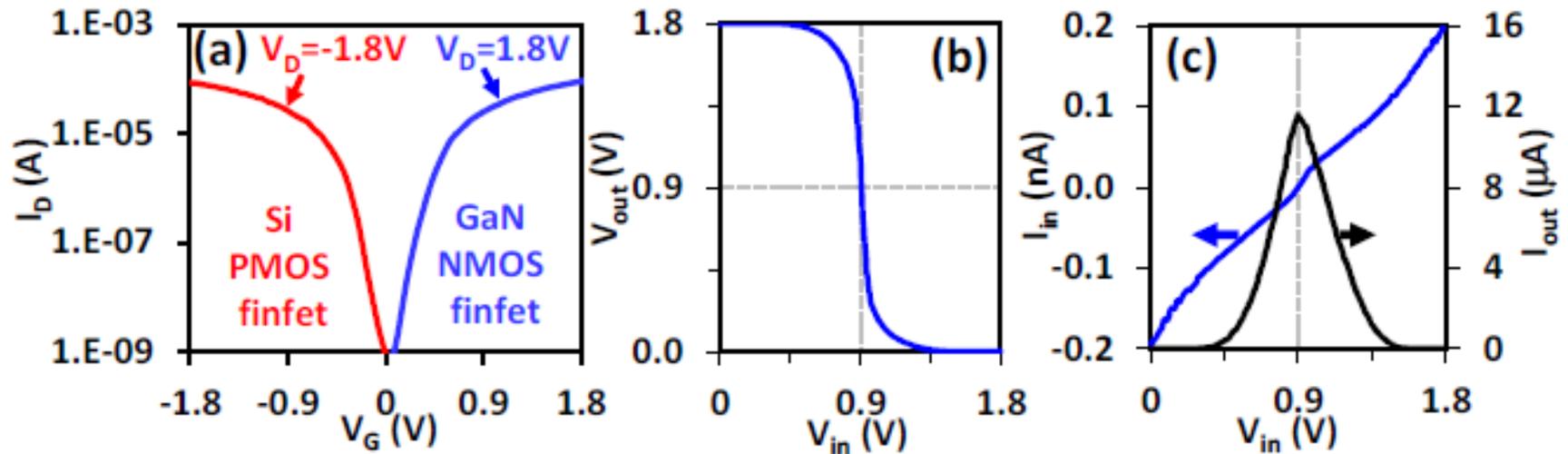
(3) 本発表の概要

- 300mm Siウェーハ上にノーマリーオフのGaNトランジスタを集積化
- ゲート長は30nmと短く、 f_{max} は400GHzに達する。さらにこの技術を進展させてSi基板上のGaNのn型FinFET上にSiのp型FinFETを3D積層してCMOSインバータ動作を実証した。
- 高耐圧・高周波GaN FETとCMOSとの融合を可能とする技術である。



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-5. 3D積層GaN-Silicon CMOS技術 (2/2)



(4)技術的に特筆すべき点

パワーIC向けに300mmSiウェーハを用いn型GaN FinFETの上にp型Si FinFETを三次元集積している点。これでパワートランジスタとCMOS集積回路の集積が可能となる。

(5)本発表が注目される理由

GaNトランジスタの性能が優れている上に、Si FinFETを集積してCMOS構成を可能としていることは素晴らしい

(6)実用化までの課題

プロセスコスト、歩留など

5-6. 反強誘電性容量を使ったFeRAM (1/2)

33.2. FeRAM using Anti-ferroelectric Capacitors (Intel)

(1) 本発表が応用される分野、製品

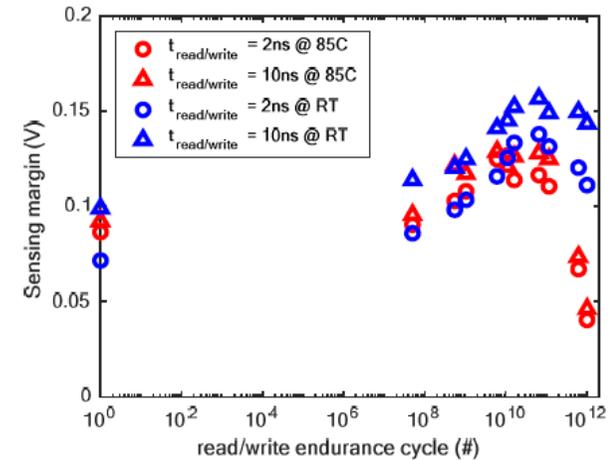
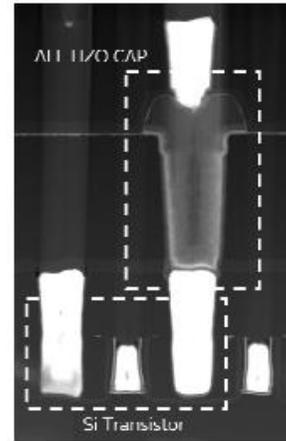
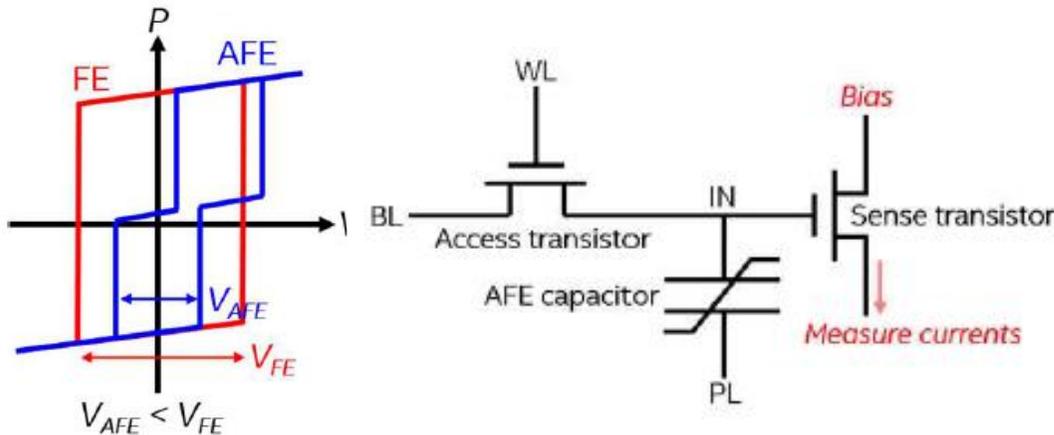
混載DRAM置き換え

(2) 当該分野における技術ニーズ

高集積密度、高速、長時間のデータ保持、高書き換え回数耐性

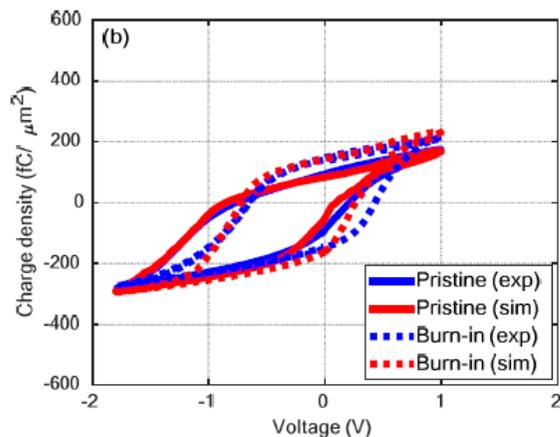
(3) 本発表の概要

- FeRAMは低電圧動作の高集積メモリとしてeDRAMの置き換えが狙える。
- 本論文では、従来の強誘電性(FE)ではなく、より低電圧動作が可能な反強誘電性(AFE)を用いる。また、1トランジスタ(1T)のFeFETではなくトランジスタと容量を分離した1T-1Cの構造をとることを提案している。
- 300mmウェーハ上にトレンチ構造を用いた微細なAFE容量を作製しており、電圧は-1.6V/1.2Vと低い。読みだし/書き込みスピードは2nsと速く、 10^{12} 回のEnduranceも達成し、より実用的なFeRAMを実現している。



(出典) 2021 IEDM(URL <https://www.ieee-iedm.org/>)

5-6. 反強誘電性容量を使ったFeRAM (2/2)



	Material	Structure	Speed (ns)	Voltage (V)	Size (μm^2)	Variation	Endurance
This work	AFE HZO	3D	2	-1.6/1.2	0.008	4σ	10^{12} at 85C
Ref. [7]	FE HZO	Planar	30	-4/4	0.1	No data	10^{11} at RT
Ref. [8]	FE HZO	Planar	14	-2.5/2.5	1	6σ	10^9 at RT

(4) 技術的に特筆すべき点

メモリ面積が極めて小さい上に、電圧は-1.6/1.2Vと低く、2nsという高速性も有している点。書き換え回数も10の12乗回を達成している点。

(5) 本発表が注目される理由

インテルが混載メモリの開発に力を入れ、上記のようなよい性能を達成しているため

(6) 実用化までの課題

集積度を上げた時の性能、歩留など

2-1-9. ISSCC 2022

目次

1. ISSCC 2022 開催概要

2. プログラム概要

3. 注目発表

3-1. プロセッサ発表概要

3-1-1. A 16nm 785GMACs/J 784-Core Digital Signal Processor Array with a Multilayers Switch Box Interconnect, Assembled as a 2×2 Dielet with 10μm-Pitch Inter-Dielet I/O for Runtime Multi-Program Reconfiguration: UCLA

3-2. AIチップ発表概要

3-2-1. 【デジタル/データフロー特化】 A 7nm Dataflow Architecture to Accelerate Software 2.0 : SambaNova

3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST

3-3. 量子コンピュータ向け極低温動作回路(クライオCMOS)発表概要

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM

1. ISSCC2022 開催概要

ISSCC(International Solid-State Circuits Conference)は、米国電気電子学会(IEEE)が主催する、**集積回路設計分野における最高峰の国際会議**として知られている。半導体のオリンピックとも呼ばれる。

- 開催時期: 2022/2/20-24 (オンライン開催)
- 学会テーマ
“Intelligent Silicon for a Sustainable World”
- 参加人数
今回の参加者数は公表されていないが、昨年(オンライン開催)が約3,000人だった為、今回も同程度と推測
- 論文データ
 - 投稿論文数: 651 / 採択論文数: 208 (採択率: 32%)
 - 地域・機関別の採択論文数

	地域別			機関別		
	北米	欧州	アジア	Industry	Academia	Joint
論文数	80	29	99	67	117	27
割合	38%	14%	48%	32%	56%	12%

- ✓ 発表はアジアが最多
- ✓ 発表の過半数は大学/研究機関

2. ISSCC2022 プログラム概要(1/3)

● 基調講演

例年、米・欧・アジアからそれぞれ1名 + 地域問わず1名 の4件



Catalysts of the Impossible: Silicon, Software, and Smarts for the Era of SysMoore

Aart de Geus, Chairman and co-CEO,
Synopsys (米国) 回路設計環境



The Future of the High Performance Semiconductor Industry and Design

Renée James, Founder, Chairman, and CEO,
Ampere Computing (米国)
Intel元幹部が創業したベンチャー企業
データセンター向けプロセッサを開発



Intelligent Sensing: Enabling the Next "Automation Age"

Marco Cassis, President, Analog, MEMS and
Sensors Group (欧州) センサー
Head of STMicroelectronics' Strategy, System
Research and Applications, Innovation Office



The Art of Scaling: Distributed and Connected to Sustain the Golden Age of Computation

Inyup Kang, President,
Samsung Electronics (韓国)
モバイル向けSoC(AI,3次元実装,無線通信)

● チュートリアル

各サブコミッティー(後述)から1名が講師となり、特定のテーマについて講義(全12テーマ)

(テーマ抜粋)

T2: Fundamentals of High-Frequency DC-DC Converters ... 電源回路について

T6: Wireless Power Transfer and Management for Medical Applications ... 無線給電回路について

T10: Fundamentals of mm-Wave Phased-Arrays ... ミリ波(無線)回路について

T12: Advances in Digital vs. Analog AI Accelerators ... AIチップについて

2. ISSCC2022 プログラム概要(2/3)

- フォーラム・ショートコース

テーマに沿って、8名程度の講師が連続して講演。今回は、フォーラムが6件、ショートコースが1件

(フォーラム)

F1: Compute-in-X (CiX): Overcoming the Data Bottleneck in AI Processing .. AI

F2: Chip Design for Low-Power, Robust, and Secure IoT Devices .. IoT

F3: The Path to 6G: Architectures, Circuits, Technologies for Sub-THz Communications, Sensing and Imaging .. 無線通信

F4: Paving the Way to 200Gb/s Transceivers .. 有線通信

F5: How to Improve AI Efficiency Further: New Devices, Architectures and Algorithms .. AI

F6: Computer Systems Under Attack – Paying the Performance Price for Protection .. セキュリティ

(ショートコース)

SC1: High Speed/High Performance Data Converters: Metrics, Architectures, and Emerging Topics .. アナログ・デジタル変換

2. ISSCC2022 プログラム概要(3/3)

● 講演

34セッション、各サブコミッティーが受け持つ(2～3セッション/サブコミッティー)

<サブコミッティー 一覧>

AN: Analog ... アナログ回路(アンプなど)全般

PM: Power Management ... 電源回路

DC: Data Converters ... デジタル-アナログ変換回路

DAS: Digital Architectures & Systems ... デジタル回路・システム

DCT: Digital Circuits ... デジタル回路・要素技術

IMMD: Image Sensor & Medical Devices ... イメージセンサ

MEM: Memory ... メモリ

RF: Radio Frequency ... 無線通信向け(ミリ波)回路

TD: Technology Directions ... 将来技術(量子コンピュータ向け回路など)

WLS: Wireless ... 無線通信向け回路

WLN: Wireline ... 有線通信向け回路

ML: Machine Learning ... 機械学習アクセラレータ回路(AIチップ)

デジタル回路・要素技術のセッションから、プロセッサに関する発表を、将来技術（量子コンピュータ向け回路など）のセッションから極低温動作回路（クライオCMOS）に関する発表を、機械学習アクセラレータ（AIチップ）のセッションからAIチップに関する発表をそれぞれ発表概要を解説、およびそのなかから注目発表を詳解する。

3. 注目発表

3-1. プロセッサ発表概要 (1/3)

ISSCCでは、基調講演の後の最初のセッションに、プロセッサに関するセッションが組まれることが通例となっている。本セッションでは発表の大半が企業で、各社の最先端プロセッサが発表される。

- 本年の発表動向

7件の講演のうち、データセンター(DC)向けが4件、PC向けが1件、モバイル向けが2件であった。このうち、5件は、**複数のチップでプロセッサを構成する、所謂チップレット構造**となっている。

(発表一覧)

2.1 Ponte Vecchio (Intel) .. **DC向けGPU 多チップ構成**

2.2 Sapphire Rapids (intel) .. **DC向けCPU 4チップ構成**

2.3 Telum (IBM) .. **DC向けCPU 2チップ構成**

2.4 Power10 (IBM) .. **DC向けCPU**

2.5 SoC (MediaTek) .. **Mobile向けSoC**

2.6 **DSP array for 5G (カリフォルニア大) .. Mobile向けDSP 4チップ構成** (注目発表として詳細を後述)

2.7 Zen3 (AMD) .. **PC向けCPU 2-9チップ構成**(チップ数を変えることでlaptopからserverまで広くカバー)

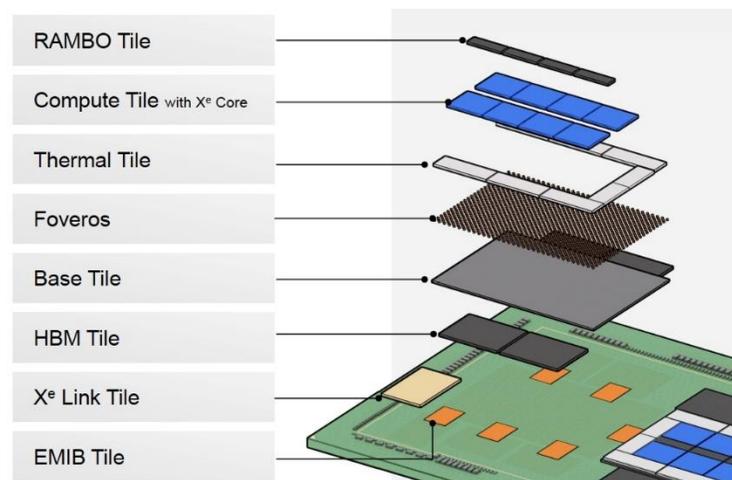
3-1. プロセッサ発表概要 (2/3)

● Intelの発表概要

Intelからは2件の発表。チップ自体は、2021年8月に開催された学会HotChipsなどで発表されていた為、既に知られた内容も多い。

2.1 Ponte Vecchio

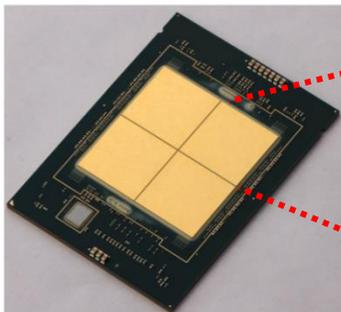
サーバ(DC)向けGPU。Nvidia A100の対抗製品。47個のタイルと呼ばれる機能チップで構成される点が特徴。プロセッサのコアは、TSMC 5nm(N5)プロセスで製造される。



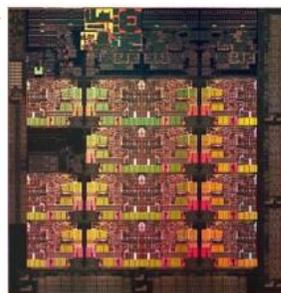
2.2 Sapphire Rapids

サーバ(DC)向けCPU(Xeon系列)。4チップで構成される点が特徴。Intel 7nmプロセスで製造される。

パッケージ写真



チップ写真



3-1. プロセッサ発表概要 (3/3)

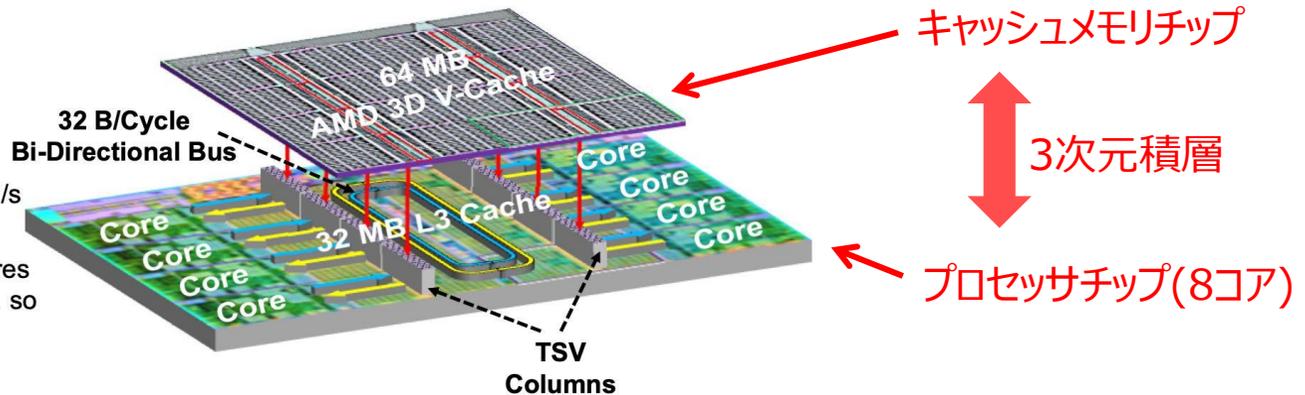
● AMDの発表概要

AMDからは1件の発表。こちらも、2021年8月の学会HotChipsで発表済み。AMDは、複数チップ構成(チップレット構造)の製品化で先行している。

2.7 ZEN3

PC/サーバ向けCPU。1-8個のプロセッサチップ(最先端プロセス)と1個のIO用チップ(前世代プロセス)で構成。プロセッサチップの個数を変えることで、ノートPCからサーバ向けまで幅広く対応が可能。キャッシュメモリを別チップで用意して、プロセッサチップ上に3次元積層している点が特徴(世界初)。これにより、従来に比べ3倍のキャッシュメモリを搭載することができ、性能が向上。TSMC 7nmプロセスで製造。

- Two columns of TSVs on left/right side of the L3 cache
- AMD 3D V-Cache extends L3 Cache capacity by 64MB (3x)
- Total inter-die bandwidth: >2 TB/s
- All control and routing to the cores is implemented on the base die, so AMD 3D V-Cache can be completely focused on density



● 今後の方向性

プロセス微細化の終焉が目前に迫ってきており、複数チップ構成(チップレット構造)によって性能向上を実現する方向性が強まると考えられる。チップレット構造では、チップ間の通信帯域が重要であり、今後、帯域増加の為の2.5/3次元実装技術の開発が盛んに行われると予測される。

3-1-1. A 16nm 785GMACs/J 784-Core Digital Signal Processor Array with a Multilayers Switch Box Interconnect, Assembled as a 2×2 Dielet with 10μm-Pitch Inter-Dielet I/O for Runtime Multi-Program Reconfiguration : UCLA (1/4)

- **本発表が応用される分野、製品**

スマートフォンや自動運転車など5G通信機能が搭載される製品

- **当該分野における技術ニーズ**

5G通信の処理に必要な演算量は多いが、スマートフォンなどに搭載される事を考慮すると低電力性と低コスト性が求められる。一方で、5G通信処理には様々な計算が必要である為、汎用性があることも求められる。

- **本発表が注目される理由**

大学の発表らしい研究要素が強い発表。一方で、大学単独の論文にもかかわらず、チップレット構造であるなど、完成度は非常に高い。

- **実用化までの課題**

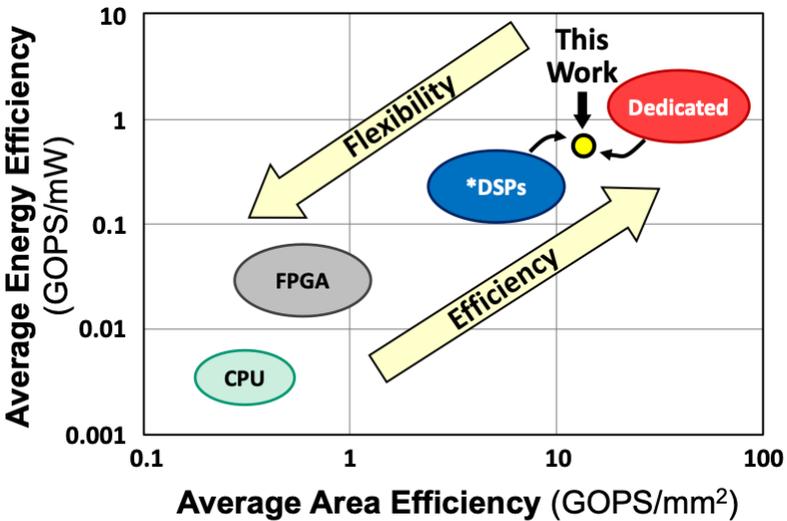
本発表で使用されるチップレット実装技術のSi-IFは、UCLAの技術であり、量産に使えるかは現時点では未知数。実用化に向けては、信頼性やコストなどが重要な課題と考えられる。

3-1-1. A 16nm 785GMACs/J 784-Core Digital Signal Processor Array with a Multilayers Switch Box Interconnect, Assembled as a 2x2 Dielet with 10μm-Pitch Inter-Dielet I/O for Runtime Multi-Program Reconfiguration : UCLA (2/4)

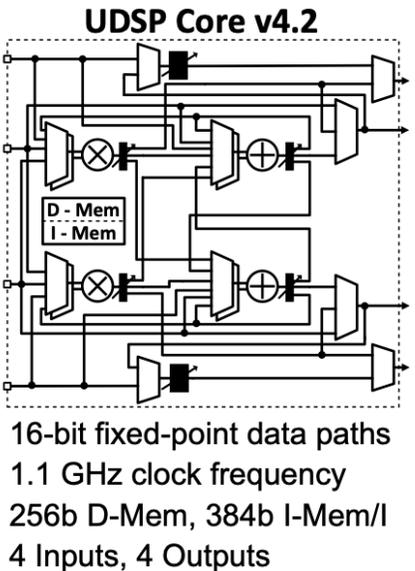
本発表の概要

- 5G通信などに使用される、信号処理専用プロセッサ(DSP)の提案
- 5G通信の処理に必要な演算量は多いが、スマートフォンなどに搭載される事を考慮すると低電力性と低コスト性が求められる。一般に、ある処理に特化した専用チップを作る事が電力・コストの両面で最善だが、5G通信処理には様々な計算が必要である為、汎用性があることも重要である(下図左参照)。
- 本提案では、信号処理の基本となるコア(UDSPと呼ばれる)を多数並べ、UDSP間の配線をつなぎ変える事で様々な信号処理を効率的に実行できる(FPGAのような)回路を提案(下図右参照)。これにより、汎用DSPと専用チップの中間の低電力性と低コスト性を旨す(下図左参照)。

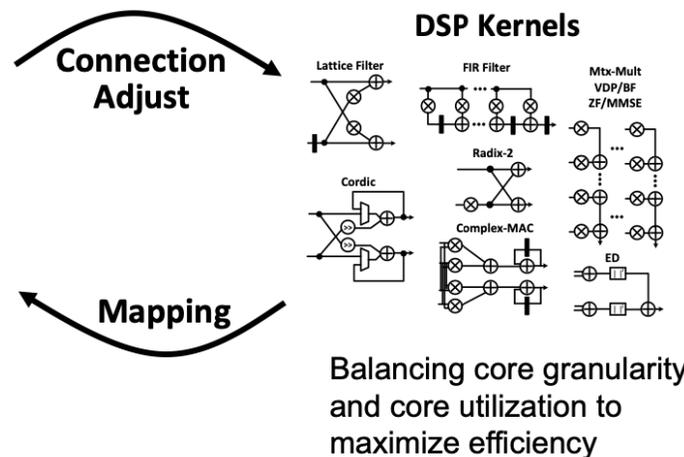
エネルギー効率(=低電力性)



面積効率(=低コスト性)



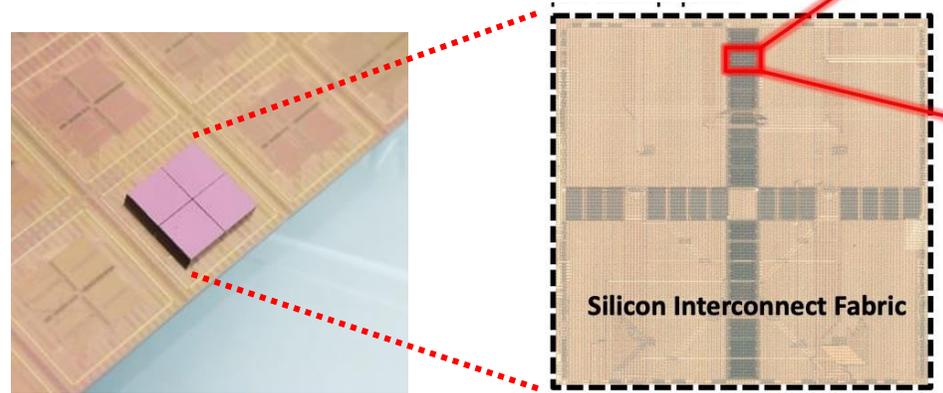
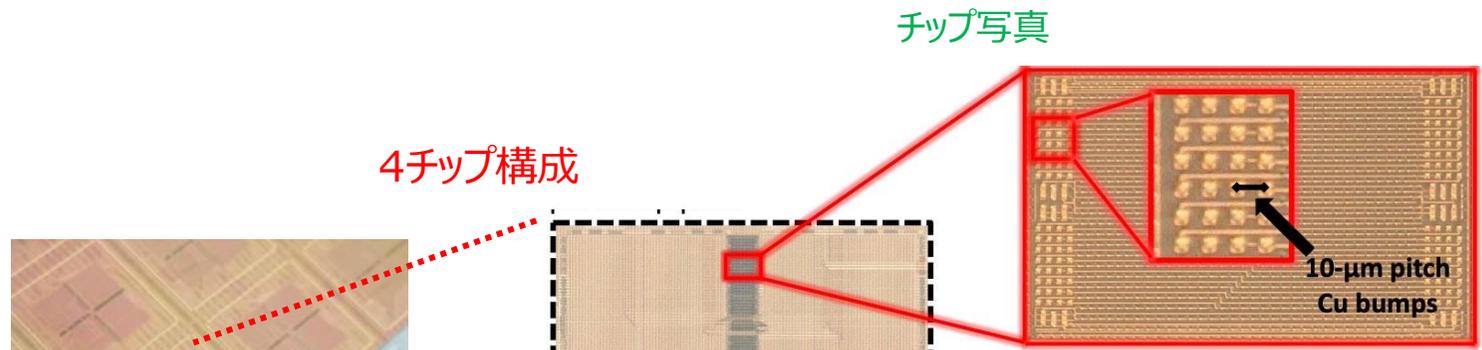
複数のUDSPのつなぎ方を変える事で
様々な信号処理を実行可能



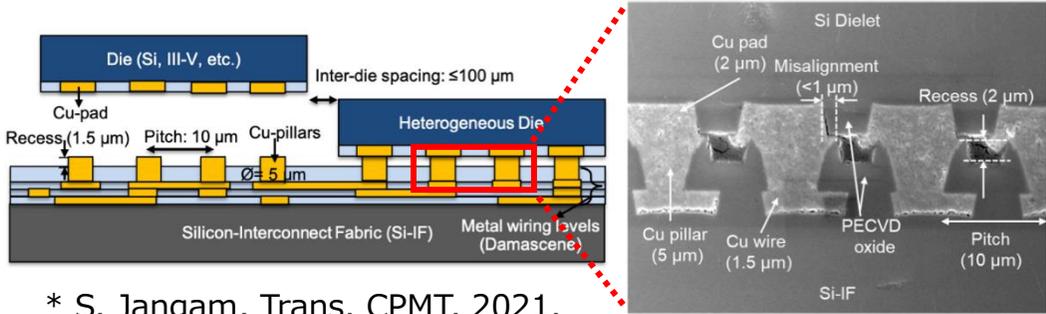
3-1-1. A 16nm 785GMACs/J 784-Core Digital Signal Processor Array with a Multilayers Switch Box Interconnect, Assembled as a 2x2 Dielet with 10μm-Pitch Inter-Dielet I/O for Runtime Multi-Program Reconfiguration : UCLA (3/4)

・ 技術的に特筆すべき点

- 信号処理コア(UDSP)を多数搭載する程、多くの機能・演算を実行できる。そこで、相互接続可能な4チップ構成を提案。チップ間通信は、UCLAが開発している、Si-IF(Silicon Interconnect Fabric)を使用。本技術により、チップ間をつなぐコネクタ(バンプ)のピッチは10mm、チップ間距離は100mm と非常に狭ピッチ・近距離で接続が可能。
- コア数が増えると、それを繋ぐ配線の混雑度が増す(コア数をNとするとN²のオーダー)。本発表では、ほとんどの配線がコア近傍になる(混雑度はNのオーダー)技術を提案。これにより、構成するチップ数を現在の4個から増やすことも可能で、高いスケーラビリティを実現。



(参考) Si-IF(Silicon Interconnect Fabric)



* S. Jangam, Trans. CPMT, 2021.

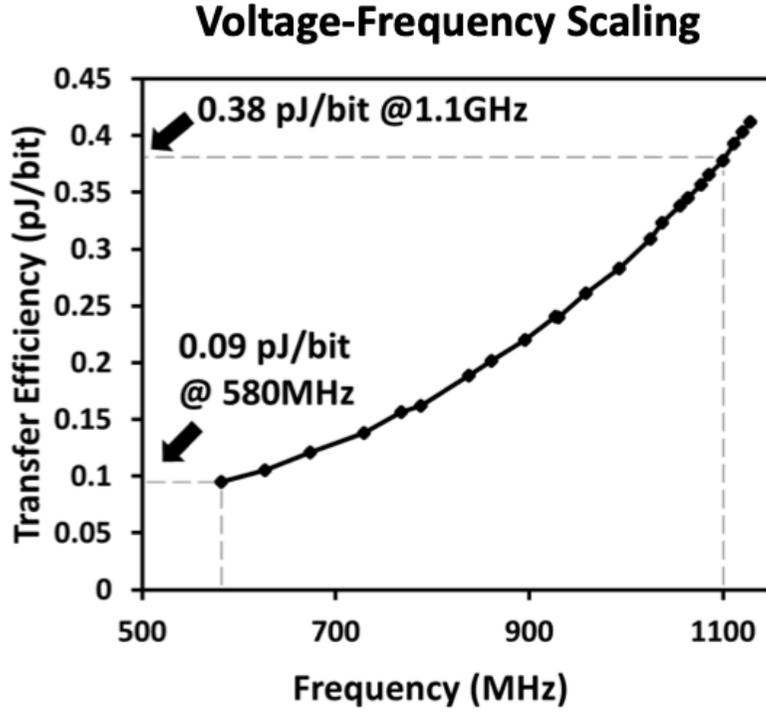
3-1-1. A 16nm 785GMACs/J 784-Core Digital Signal Processor Array with a Multilayers Switch Box Interconnect, Assembled as a 2x2 Dielet with 10μm-Pitch Inter-Dielet I/O for Runtime Multi-Program Reconfiguration : UCLA (4/4)

● 既存研究との比較

狭いバンプ(コネクタ)ピッチでチップ間を接続

Spec	This Work	CICC'21	JSSC'20	ISSCC'18
Techology	16nm	16nm	7nm	16nm
Package Substrate	2-Layer Si-IF	4-Layer EMIB	15-Layer CoWoS	MCM/PCB
Bump Pitch (μm)	10	55	40	150
Reach (μm)	350	3,000	500	80,000
Data Rate (Gbps/pin)	1.1	2	8	25
Voltage (V)	0.8	0.9	0.3	0.95
Energy Effic. (pJ/bit)	0.38	0.83	0.56	1.17
I/O Area Density (μm²/bit)	137	203	500	10,175
Peak Shoreline BW Density (Gbps/mm)	297	256	1600	292
Layer BW Density (Gbps/mm/layer)	149	64	107	25

エネルギー最小



3-2. AIチップ発表概要（1/4）

近年、深層学習(Deep Learning、以下DL)を核としたAI技術が、目覚ましい進展を遂げている。これに伴い、**必要な計算量が著しく増加しており、AI処理(主にDL)に特化した専用回路(アクセラレータ)、所謂AIチップの研究開発が盛ん**になっている。ISSCC2022でも、多数のAIチップ関連発表があった。

● 発表動向

ISSCC2022では、AIに関連する2件のフォーラム、5件のセッション(25件の発表)が行われた。

(AIに関連するフォーラム)

- F1: Compute-in-X (CiX): Overcoming the Data Bottleneck in AI Processing
- F5: How to Improve AI Efficiency Further: New Devices, Architectures ...

(AIに関連するセッション)

- S11: Compute-in-Memory and SRAM
- S15: ML Processors
- S16: Emerging Domain-Specific Digital Circuits and Systems
- S21: Highlighted Chip Releases: Digital/ML (Invited)
- S29: ML Chips for Emerging Applications

AIチップは、製品としてはGoogleやAppleなど米国企業が強い分野ではあるが、ISSCCではそのような企業からの発表は少なく、大学からの研究要素の強い発表が大半を占める状況である。また、AIチップ関連発表の半数は、後述するCIM関連の発表であり、さらにそのほとんどはTSMCと台湾・中国の大学が組んだものであった。このような背景もあり、AIチップ関連は、アジア勢(特に、台湾・中国)が非常に強い領域となっている。

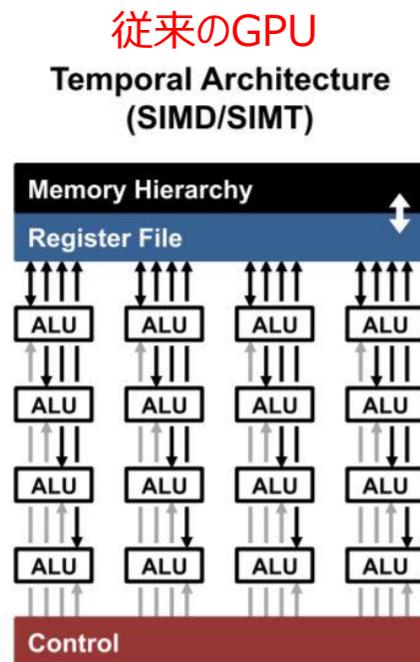
3-2. AIチップ発表概要 (2/4)

● 近年の研究動向

大きく分けて、以下の2つの方向性で研究開発が進んでいる。

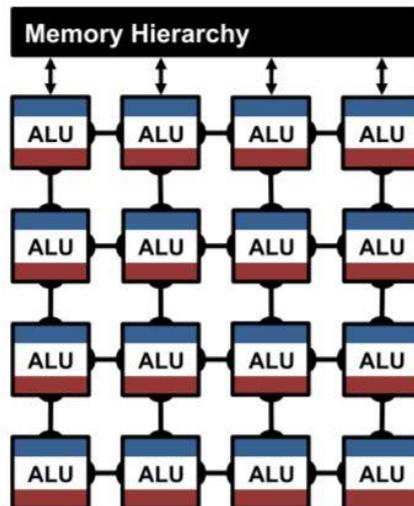
(1) デジタル(データフロー特化) → 性能重視 (DC向けAI学習、スマートフォンでの高度なAI推論など)

- DLにおける計算の多くは積和算で構成されており、従来から積和算を高速・高効率に実行できるGPU (Graphics Processing Unit)が広く利用されてきたが、近年、GPUよりも高速・高効率にDL処理を実行する技術として、DL特有のデータフローに特化し、データ転送などを最適化する技術の研究開発が進んでいる。
- 現在、GoogleやAppleなど多数の企業がAIチップを実用化しているが、これらは基本的に、このアプローチを採用している。本年のISSCCでは、ベンチャー企業2社(SambaNovaとTenstorrent)の招待講演が行われた。SambaNovaの発表を注目発表として詳細を後述する。



AIチップ(データフロー特化)

Spatial Architecture (Dataflow Processing)



DLに特徴的なデータフローを考慮した構造
→メモリと演算器間のデータ移動を削減し、
高速・高効率にDL処理を実行可能

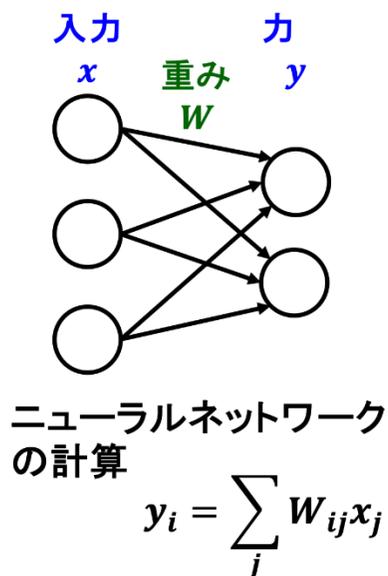
* V. Sze, Proc. IEEE, 2017.

3-2. AIチップ発表概要 (3/4)

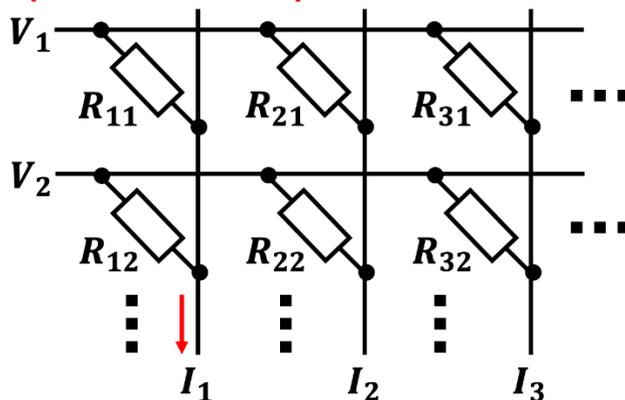
● 近年の研究動向 (つづき)

(2) アナログ(CIM; Compute-in-Memory) → 低電力重視 (IoT端末でのAI推論など)

- 現在のAIチップは、前述の通り、メモリから演算器にデータを移す必要がある。このデータ転送のエネルギーを削減する為、メモリ上で計算を実行する技術(CIM; Compute-in-Memory 又は IMC; In-memory Computing などと呼ばれる)が提案されている。本技術は、電流や電荷などを用いたアナログ計算を活用したものである。アナログ演算は低ビット精度ではデジタル計算よりも高効率に計算可能であり、結果、CIMは、高エネルギー効率でDL処理を実行可能である。(高いビット精度が必要なDL学習には向いていない)
- CIMは、近年盛んに研究開発が行われている分野で、本年もAIチップ関連の発表の半数はCIM関連であった。その中から、アナログメモリを活用したKAISTの発表を注目発表として詳細を後述する。



メモリ(可変抵抗アレイ)を用いた積和算実装 (CIM)



可変抵抗アレイをどのように実現？

- ✓ SRAM
- ✓ 新原理不揮発メモリ (ReRAM, MRAM, PCM ..)

$$I_i = \sum_j \frac{1}{R_{ij}} V_j$$

← キルヒホッフの電流則に基づいた電流によるアナログ計算

3-2. AIチップ発表概要（4/4）

- 今後の方向性

データフロー特化したデジタルAIチップは、実用化が進んでいる事もあり、研究論文としては今後減少傾向になる可能性がある。一方で、CIMは、まだ研究段階で、実用化事例も殆どなく、実用化に向けた技術開発が、今後も盛んに行われると推測される。

現在、CIMの発表の多くは、SRAMベースのものである。これは、DL処理をCIMで実行するには、ある程度の規模のメモリが必要で、現時点ではSRAMで実現する事が容易であるからである。一方で、SRAMではなくReRAMやMRAMなど新原理不揮発メモリを使ったものも、発表が増えつつある(本年のISSCCでは3件の発表があり、全てTSMCが共著に入っている)。これまで、これらのメモリ・デバイスを高い集積度でCMOS回路と混載することが難しく、ISSCCなどの集積回路設計系の学会での発表は限定的だった。今後、メモリ・デバイスの完成度、CMOSとの混載技術が向上するにつれ、発表は増加していくと予想される。

新原理不揮発メモリを利用するメリットは、(1)電源をオフにしても情報を保持可能、(2)同じ面積でSRAMよりも大容量化可能=CIMの観点では、面積当たりより多くの演算を実行可能、の2点である。

3-2-1. 【デジタル/データフロー特化】 A 7nm Dataflow Architecture to Accelerate Software 2.0: SambaNova* (1/4)

- ・ **本発表が応用される分野、製品**

スパコン、データセンター向けのAIアクセラレータ

- ・ **当該分野における技術ニーズ**

近年、自然言語処理を筆頭にDLモデルの巨大化が進んでいる。この巨大モデルの学習には、膨大な計算が必要であり、現在広く使われているGPUよりも効率的にDL処理を実行できるAIチップが求められている。

- ・ **本発表が注目される理由**

DLモデルの大規模化は今後も続くと考えられており、本提案は、その学習を効率的に実行できる(例えば、従来のGPUでは8ラック必要だったものが、提案チップを使うと0.5ラックで可能になる)ものであり、注目に値する。

*シリコンバレーのAIチップ・ベンチャー企業で、2021年4月の資金調達で評価額は50億ドル超となっている。

<https://eetimes.itmedia.co.jp/ee/articles/2106/23/news038.html>

3-2-1. 【デジタル/データフロー特化】 A 7nm Dataflow Architecture to Accelerate Software 2.0: SambaNova (2/4)

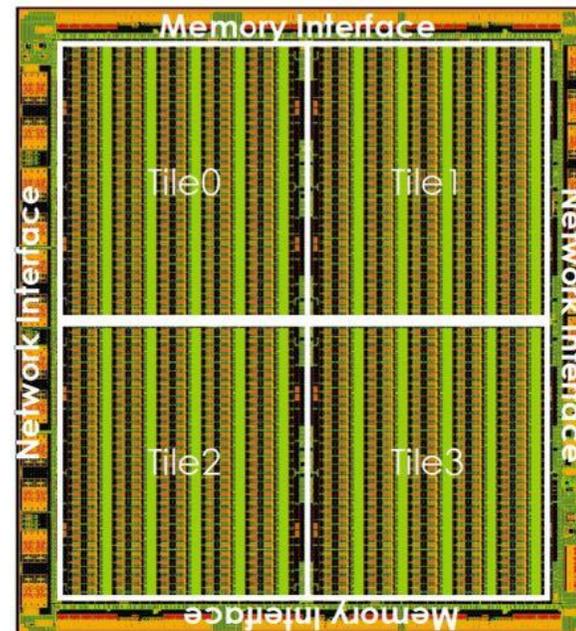
• 本発表の概要

- DL処理は、データのフローとして表すことができる。様々なデータフローに最適化可能なアーキテクチャを有するAIチップ“SN10 RDU”を提案。
- “SN10 RDU”は、TSMC 7nmプロセスで製造されている。

- **640 Pattern Compute Units (PCU)**
 - 300+ BF16 TFLOPs
- **640 Pattern Memory Units (PMU)**
 - 300+ MB on-chip memory
 - 150 TB/sec Bandwidth
- **6 DDR4 channels up to 3200 MHz**
 - 1.5 TB memory capacity
 - Up to 3200 MHz
- **Software Configurable 4 Tiles**
- **4 PCIe Gen4 channels**
- **TSMC 7nm Process**
 - 40+ billion transistors
 - 50 Km of wires

チップ写真

SN10 Reconfigurable Dataflow Unit (RDU)

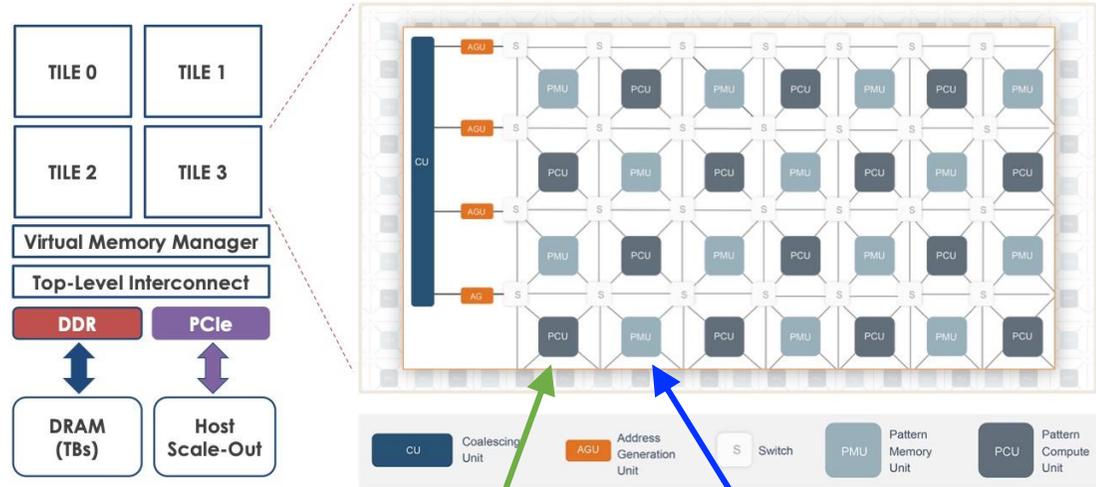


3-2-1. 【デジタル/データフロー特化】 A 7nm Dataflow Architecture to Accelerate Software 2.0: SambaNova (3/4)

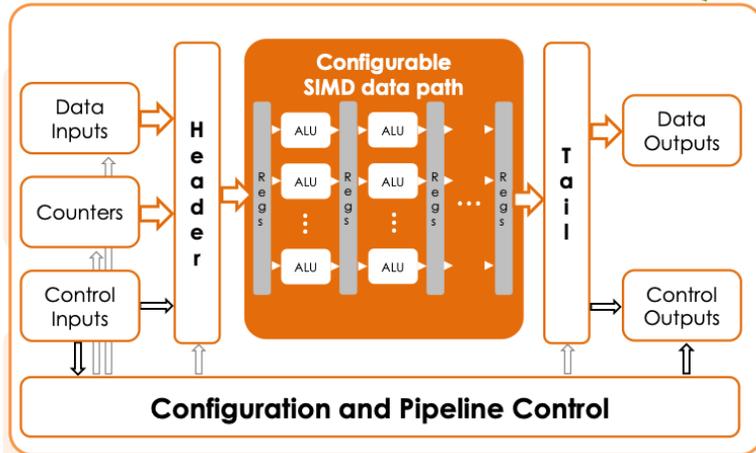
・技術的に特筆すべき点 (1/2)

“SN10 RDU”は、PCUと呼ばれる演算ユニットと、PMUと呼ばれるメモリユニットで構成される。PCUとPMUは格子状に配置。PCUは様々なデータパスを設定でき、DLの演算を効率的に実行可能

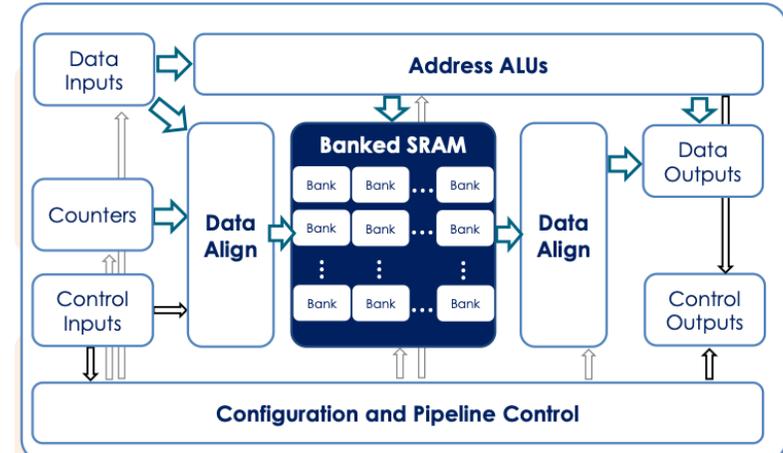
全体構成



PCU



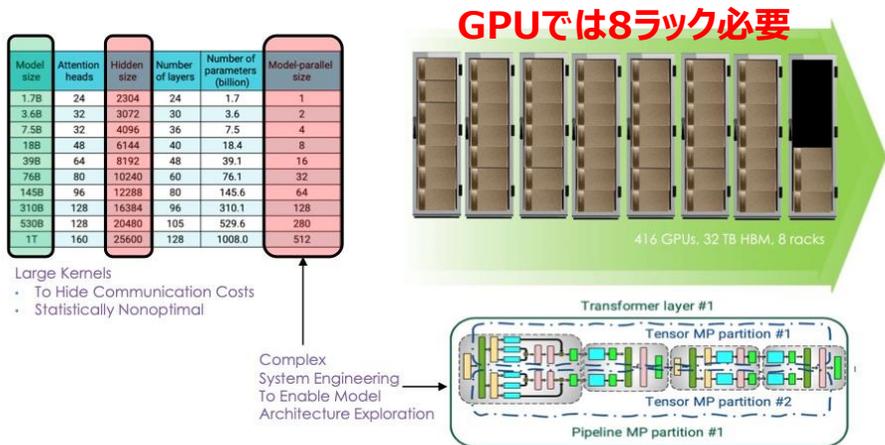
PMU



3-2-1. 【デジタル/データフロー特化】 A 7nm Dataflow Architecture to Accelerate Software 2.0: SambaNova (4/4)

・ 技術的に特筆すべき点 (2/2)

- ・ 1兆パラメータという巨大な自然言語処理モデルの学習に、従来のGPUでは8ラック必要だが、提案の“SN10 RDU” を使えば0.5ラックで可能になる。
- ・ 非常に解像度の高い(50k x 50k)画像の学習について、提案の“SN10 RDU” が最高の精度を達成



Out-of-Box Models

- Huggingface Models
- Write yours in PyTorch

Developer Efficiency

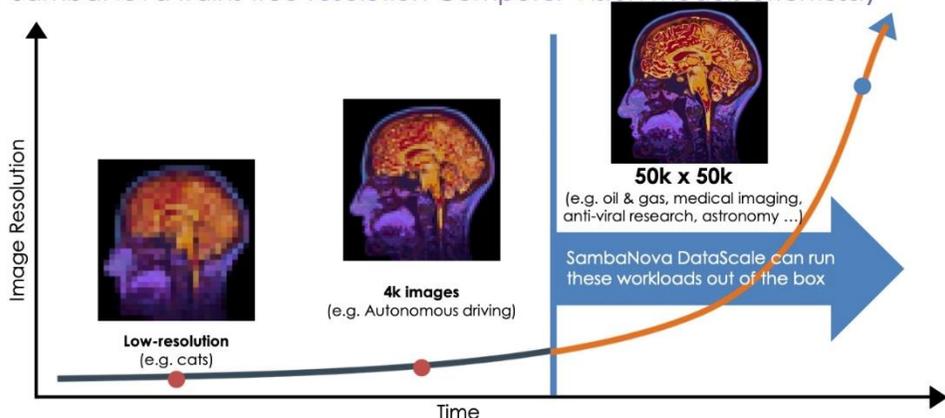
- Focus on ML problems instead of System Engineering

High Accuracy Models

- No Compromise on Model Architecture required to hide System Deficiencies

SN10 RDU (提案) では0.5ラック

SambaNova trains true-resolution Computer Vision models effortlessly



World record accuracy in high-res convolution training

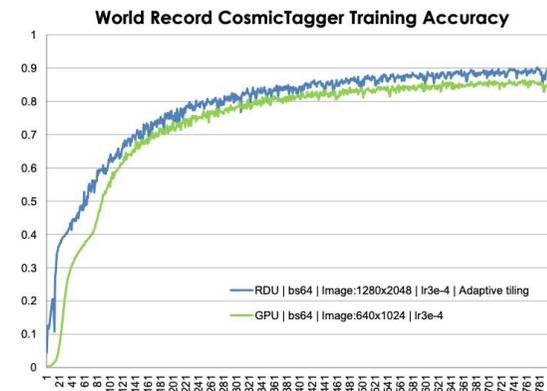
90.23% Accuracy

IEEE Xplore®
 EDITORS: Volodymyr Khutornenko, Vadymovskiy, Lucaciu
 Anna Sidor, Anna Petrova

DEPARTMENT: NOVEL ARCHITECTURES

Accelerating Scientific Applications With SambaNova Reconfigurable Dataflow Architecture

Murali Emami, Venkatesh Vishwanath, Corey Adams, Michael E. Papka, and Rick Stevens, Argonne National Laboratory, Lemont, IL, 60433, USA



3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST (1/5)

・ 本発表が応用される分野、製品

AI推論処理(学習は行わない)に高いエネルギー効率が求められる機器(スマートフォン、IoT機器など)

・ 当該分野における技術ニーズ

・ 従来、AI処理は、クラウドで実行される事が多かった。例えば、スマートフォンの音声認識は、スマートフォンのマイクで収録された音声を一括クラウドに送り、そこで音声認識の処理を実行し、その結果を端末に戻す、という一連の処理が行われる。近年、通信に伴うレイテンシやセキュリティの課題から、端末上でAI処理の実行が求められている。

・ 本発表が注目される理由

現在、製品化されているAIチップのほぼ全てはデジタル計算によるものである。AI推論処理については、低精度の演算でも実用上問題ないケースがある事が分かっており、低精度の演算でデジタルよりも高効率で計算可能なアナログ演算の導入が検討されている。一方で、アナログ演算には、温度などの環境変化や製造ばらつきに弱いなど、幾つか課題があり、その解決策を提示している本論文は、実用化に向けた技術として興味深い。

・ 実用化までの課題

提案の回路は、適用できるDLモデルに制限がある。より汎用性をもった回路の実現が今後の課題である。

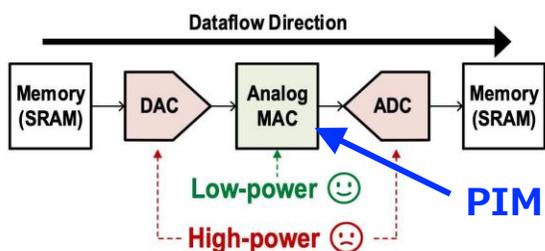
3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST (2/5)

本発表の概要 (1/2)

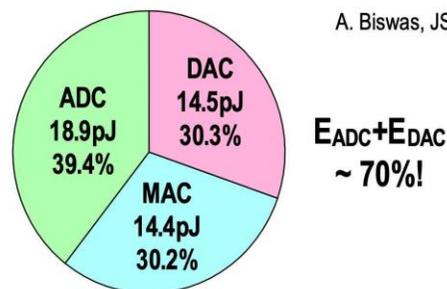
- アナログ演算によるDL処理アクセラレータを提案。アナログ演算には、① 演算の前後にアナログ・デジタル間の変換が必要で、その電力オーバーヘッドが大きい、② 温度・電圧変化や製造ばらつきに弱い、という課題があった。
- 本発表では、課題①に対して、アナログメモリを使用する事を提案し、課題②については、後述のばらつきの影響を低減する技術を提案。

課題① アナログ・デジタル変換(DAC/ADC)のオーバーヘッドが大きい

Dataflow of the conventional ML processors



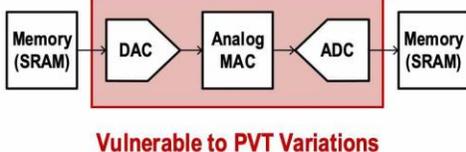
Energy breakdown of the operation [8]



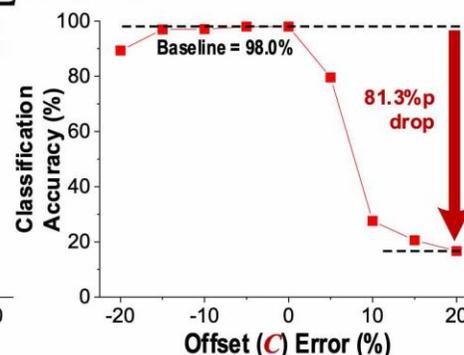
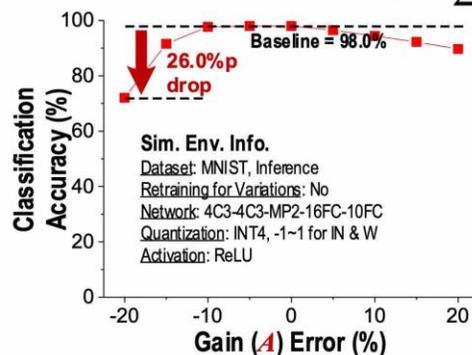
A. Biswas, JSSC 2019

課題② 温度・電圧変化や製造ばらつきに弱い

Dataflow of analog ML processor



$$Y = A \sum IN \times W + C$$

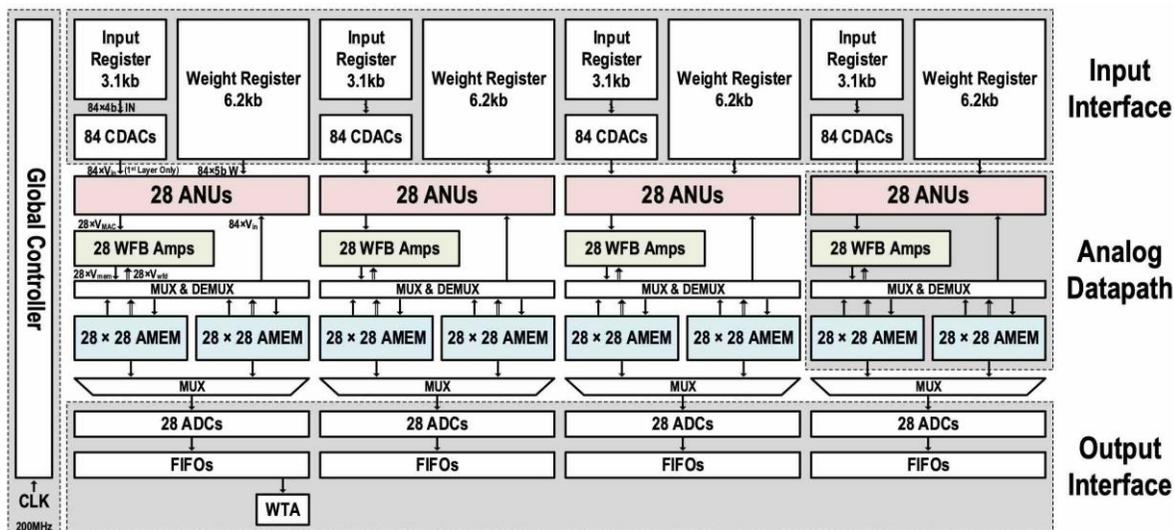


Sim. Env. Info.
 Dataset: MNIST, Inference
 Retraining for Variations: No
 Network: 4C3-4C3-MP2-16FC-10FC
 Quantization: INT4, -1~1 for IN & W
 Activation: ReLU

3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST (3/5)

本発表の概要 (2/2)

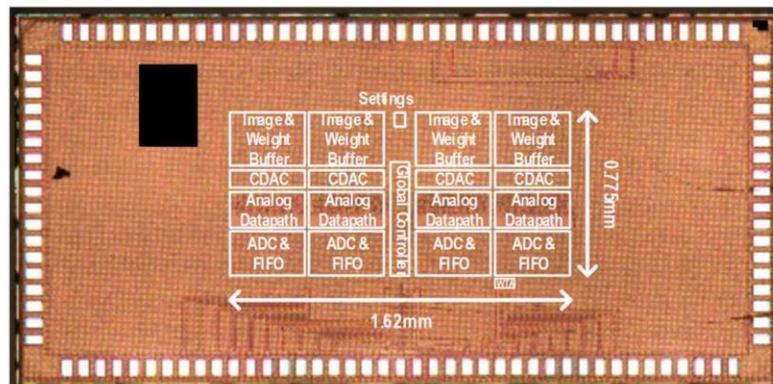
- 28nmプロセスで試作し、100TOPS/Wを超える、非常に高いエネルギー効率を達成



← 回路全体構成

ANU: アナログ演算ユニット(CIM)
AMEM: アナログメモリ

チップ写真 →
(28nm CMOS)



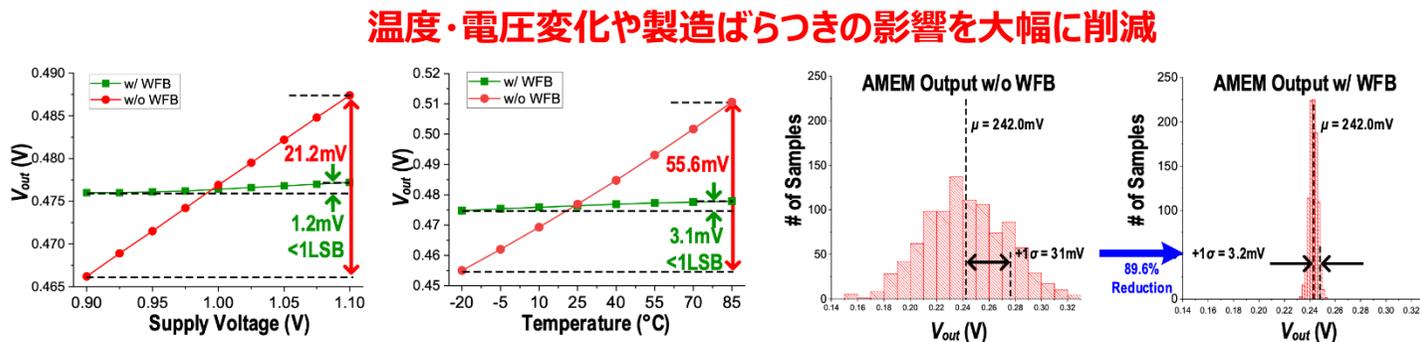
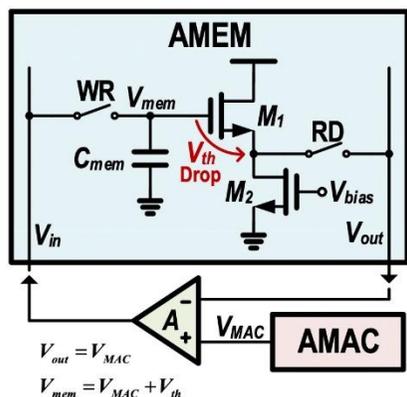
Technology (nm)	28
Bit Precision (b)	5 (weight) / analog (input)
Power (Datapath, μ W)	182 (peak) 303 (average) 604 (worst)
Energy/OP (Datapath, fJ/OP)	0.072 (peak) 0.120 (average) 0.24 (worst)
Energy Efficiency (Datapath, TOPS/W)	552.5 (peak) 332.7 (average) 166.8 (worst)
Full Processing w/o Data Conversion	○
Variation Tolerance	○

3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST (4/5)

・技術的に特筆すべき点 (1/2)

課題② アナログ演算が温度・電圧変化や製造ばらつきに弱い、という課題への対策として、(a) メモリ書き込みのフィードバック技術、(b) 計算に必要なパルス幅のオフセット除去技術 を提案。

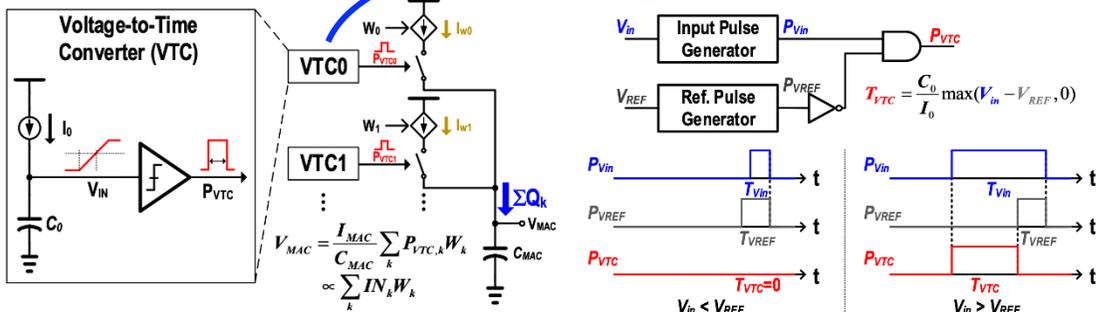
(a) メモリ書き込みのフィードバック技術



(b) パルス幅のオフセット除去技術

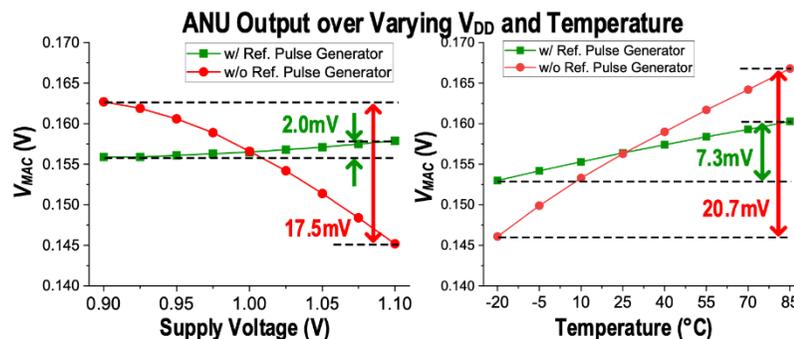
パルス幅を計算に使用

オフセット除去回路(提案)



Block diagram of MAC operation in ANU

温度・電圧変化の影響を大幅に削減



3-2-2. 【アナログ/CIM】 A 332.7TOPS/W 5b Variation-Tolerant Analog CNN Processor Featuring Analog Neuronal Computation Unit and Analog Memory: KAIST (5/5)

・ 技術的に特筆すべき点 (2/2)

アナログ演算により、非常に高いエネルギー効率を達成

	[1] ISSCC 2021	[2] JSSC 2019	[3] VLSI 2016	[4] JSSC 2019	[5] JSSC 2017	[6] JSSC 2019	This Work
Technology (nm)	65	28	65	28	130	55	28
Computing Method	In-memory Computing	VCO-Counter	DSM	Charge Redistribution	Charge Redistribution	VCO-Counter	Time-domain & Charge Accum.
Area (mm ²)	3.3 (System)	960	0.9504	5.76 (System)	0.012	3.4 (System)	1.26 (System)
Clock Frequency (MHz)	105	753	0.1	-	2500	-	200
Bit Precision (bit)	4, 8 (weight) 2, 4, 8 (input)	8	16	1	4 (weight) / 6 (input)	6	5 (weight) / analog (input)
Throughput (GOPs) ¹⁾	1720	2.06	3.23	478	5.00	2.15	100.8
Power (System ²⁾ / Datapath ³⁾ (μW)	38400 ^{a)} / 4060	N. A. / 166	N. A. / 3899 ^{d)}	899 ^{d)} / 583 ^{d)}	647 ^{b) d)} / -	690 (peak) / -	4637 / 182 (peak) 5073 / 303 (mean) 5375 / 604 (worst)
Energy Efficiency (System / Datapath) ⁴⁾ (TOPS/W)	28.6 ^{a) d)} / 138.3 ^{d)}	N. A. / 31.74 ^{b) d)}	N. A. / 8.49 ^{b) d)}	21.3 ^{d)} / 32.80 ^{d)}	7.41 ^{b) d)} / -	4.49 (peak) / -	21.7 (peak) ^{c)} / 552.5 (peak) ^{c)} 19.9 (mean) ^{c)} / 332.7 (mean) ^{c)} 18.8 (worst) ^{c)} / 166.8 (worst) ^{c)}
Energy per Operation (Datapath) (fJ/op) ⁴⁾	0.288 ^{d)}	1.26 ^{b) d)}	4.71 ^{b) d)}	1.22 ^{d)}	-	-	0.072 (peak) ^{c)} 0.120 (mean) ^{c)} 0.24 (worst) ^{c)}
Full Processing w/o Data Conversion	X	X	X	X	X	X	0
PVT Tolerance	X	X	X	X	X	X	0

1) 1 OP = 1 MUL or 1 ADD. (1 MAC = 1 MUL + 1 ADD = 2 OPs)

2) Total power consumption including the power consumption of datapath, digital logics and on-chip memories.

3) Including the power consumption of MAC and on-chip memory for feature map or weight.

4) Energy efficiency normalized for 5-b weight and 5-b input.

a) Does not include the power consumption of feature map transfer from/to off-chip memory.

b) Does not include the power consumption of feature map memory read/write.

c) Calculated with effective number of input bits = 5.

d) Mean performance at the nominal condition

温度・電圧変化や製造ばらつきに弱い、
という課題への対策を実施

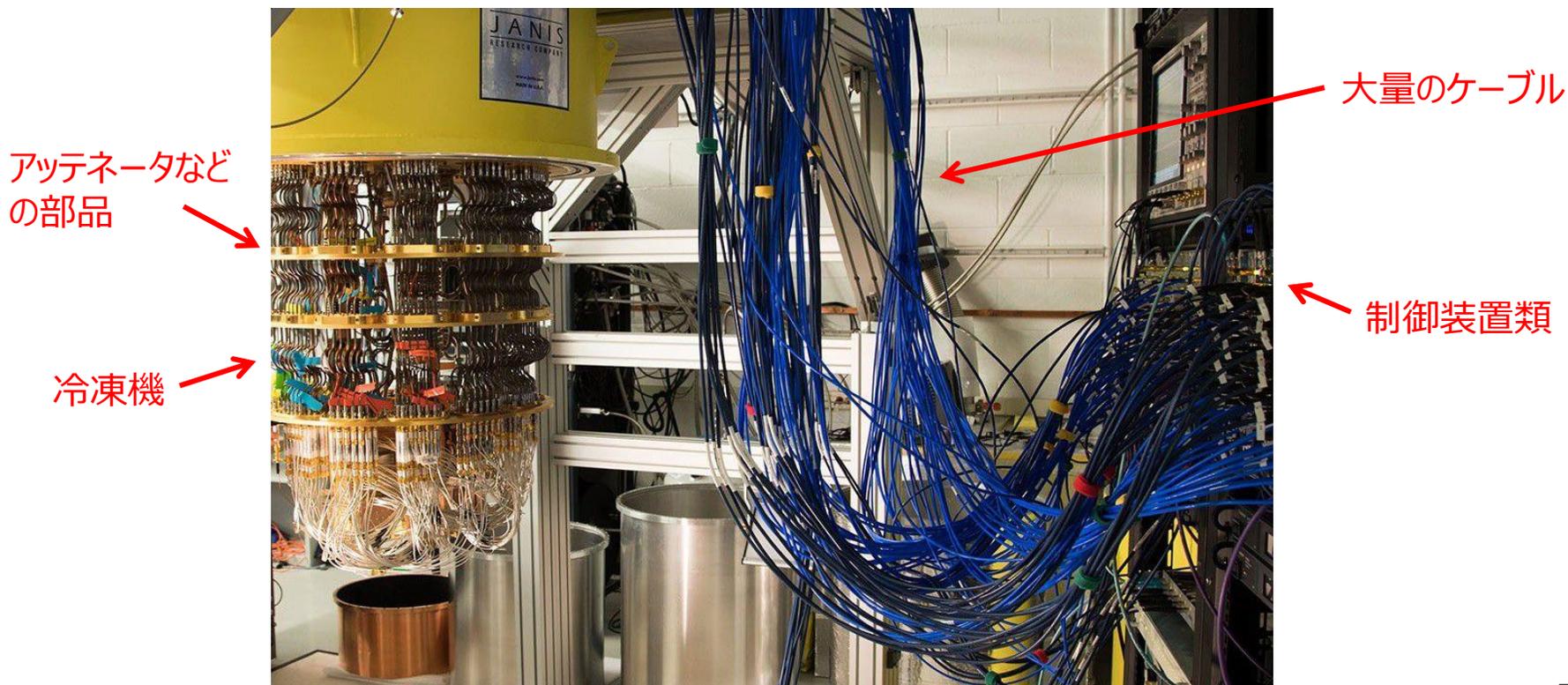
高いエネルギー効率を達成

3-3. 量子コンピュータ向け極低温動作回路(クライオCMOS)発表概要 (1/4)

量子コンピュータ向けの4K以下の極低温で動作する集積回路(一般的にクライオCMOSと呼ばれる)の研究開発が近年盛んになっており、ISSCC2022では、一般講演3件、招待講演2件の関連発表があった。

● 研究の背景

近年、超伝導や半導体を用いた量子コンピュータの研究開発が盛んに行われている。これらのコンピュータは下図のように、冷凍機内の量子ビットと室温に設置された制御装置類が、ケーブルとアッテネータなどの部品で接続された構成となっている。現状では、1量子ビットに対して複数のケーブルが必要であり、量子ビットの増加に伴い、ケーブル数や部品点数の削減などの対策が望まれている。



3-3. 量子コンピュータ向け極低温動作回路(クライオCMOS)発表概要 (2/4)

● 研究の背景 (つづき)

前述の通り、量子ビットの制御には多数の信号線が必要である。量子ビットが数個程度の場合は、汎用の測定器で対応ができたが、量子ビットの増加に伴い、市販の汎用部品を組み合わせた専用ハードウェアが使用されるようになってきている。今後、量子ビットが数千・数万となった際は、これらのハードウェア機能(の一部を)を半導体チップに集積、さらに冷凍機内の量子ビット近傍に設置する事が検討されている(所謂クライオCMOS)。これにより、冷凍機を出入りするケーブル数・部品点数を削減できることが期待されている。

現在

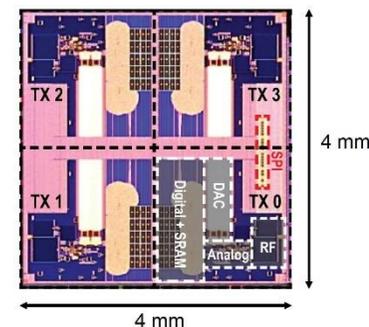
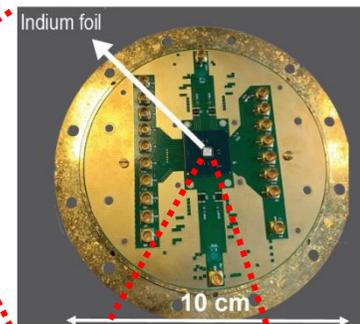
将来 (量子ビット > ~1,000)



Chip at 3 K plate

60 nm

Qubit at 10 mK



専用の制御ハードウェア
(汎用部品の組み合わせ)

1チップに集積

* B. Patra, ISSCC, 2020.

3-3. 量子コンピュータ向け極低温動作回路(クライオCMOS)発表概要 (3/4)

● ISSCCにおける発表動向

Intel/Delft工科大学連合が、クライオCMOS分野の研究開発をリードしている。その他、GoogleやIBM、LETIなど量子コンピュータを開発している企業・研究機関もISSCCで発表している。

2018 Delft工科大学がクライオCMOSに関して初めて発表 .. [要素回路レベル](#)

2019 Googleが量子ビット制御チップを発表 .. [1量子ビットの制御機能を実装](#)

2020 Intelが量子ビット制御チップ(HorseRidge)を発表 .. [複数量子ビット同時制御機能を実装](#)

2021 Intelが後継のクライオCMOSチップ(HorseRidge2)を発表 .. [量子ビットの状態読み出し機能を実装](#)

2022 下記の一般講演3件、招待講演2件

(一般講演)

- IBM .. 量子ビットの制御チップを発表。 ([注目発表として詳細を後述](#))
- POSTECH(韓国) .. 量子ビットの制御・読み出しチップを発表。POSTECHは最近活発に発表している。
- EFPL(スイス) .. 量子ビット制御回路に使われる発振器を開発。研究を主導するE. Charbon教授は、2016年までDelft工科大学でも教授を務め、ISSCCにおける初のクライオCMOS発表(2018年)も主導した。

(招待講演)

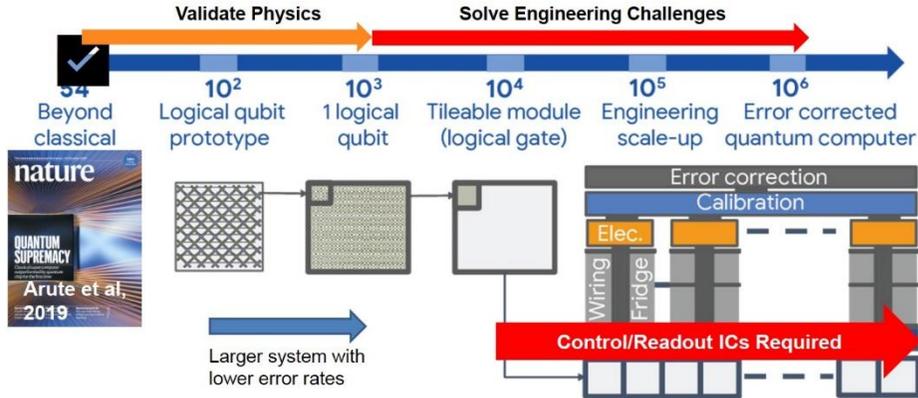
- IBM)
- Google)

3-3. 量子コンピュータ向け極低温動作回路(クライオCMOS)発表概要 (4/4)

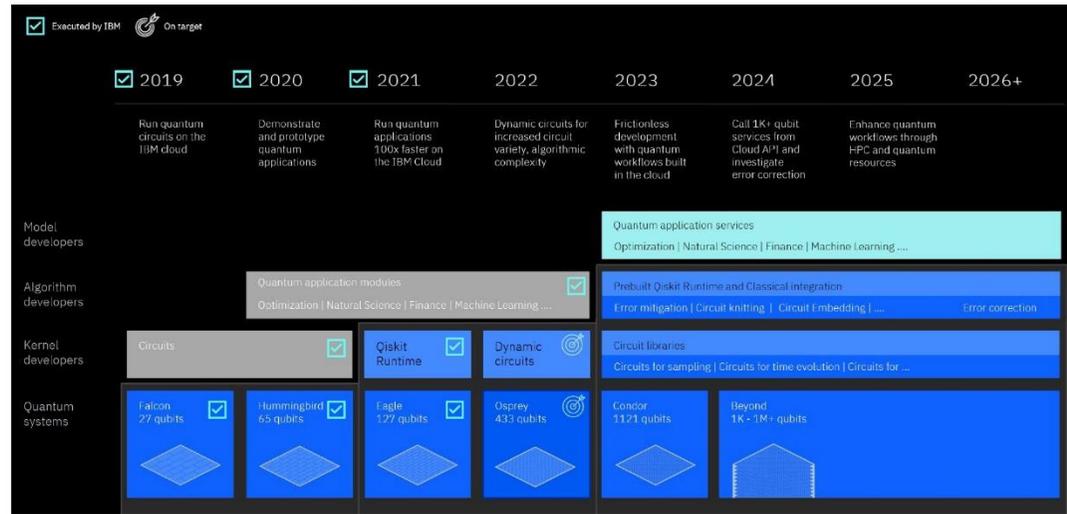
● 将来の方向性

クライオCMOSは、量子ビット数が現在よりも相当に増えた際に必要となる技術である。IBMとGoogleの招待講演によると、2,000~10,000量子ビット以上でクライオCMOSが必要になる、という見解のようである。IBMのロードマップを踏まえると、最速で2024年頃に実際の量子コンピュータに使われる可能性がある。

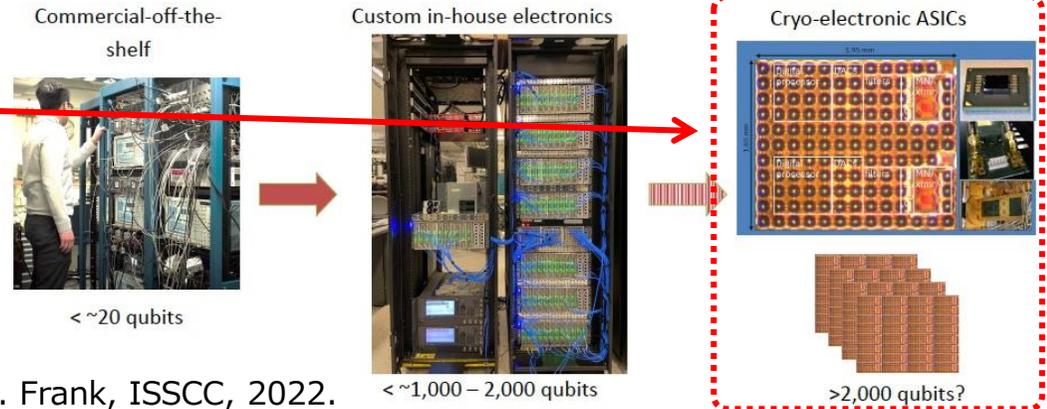
Googleのロードマップ^o (J. Bardin, ISSCC2022)



IBMのロードマップ^o (G. Zettles, ISSCC2022)



2,000量子ビット以上で
クライオCMOSチップが
必要になる



* D.J. Frank, ISSCC, 2022.

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM (1/5)

- **本発表が応用される分野、製品**

超伝導量子コンピュータ (技術的には半導体スピン量子コンピュータにも適用可能)

- **当該分野における技術ニーズ**

現在の超伝導量子コンピュータは室温の制御装置が用いられている(p.26-27参照)。今後、量子ビット数の増加に伴い、装置と量子ビットを繋ぐケーブル数・部品点数の削減が望まれており、その解決策として、現在室温に置かれている制御装置の機能(の一部)を、半導体チップに集積し、量子ビット近傍の極低温下で動作させる技術(所謂クライオCMOS)に期待が高まっている。

- **本発表が注目される理由**

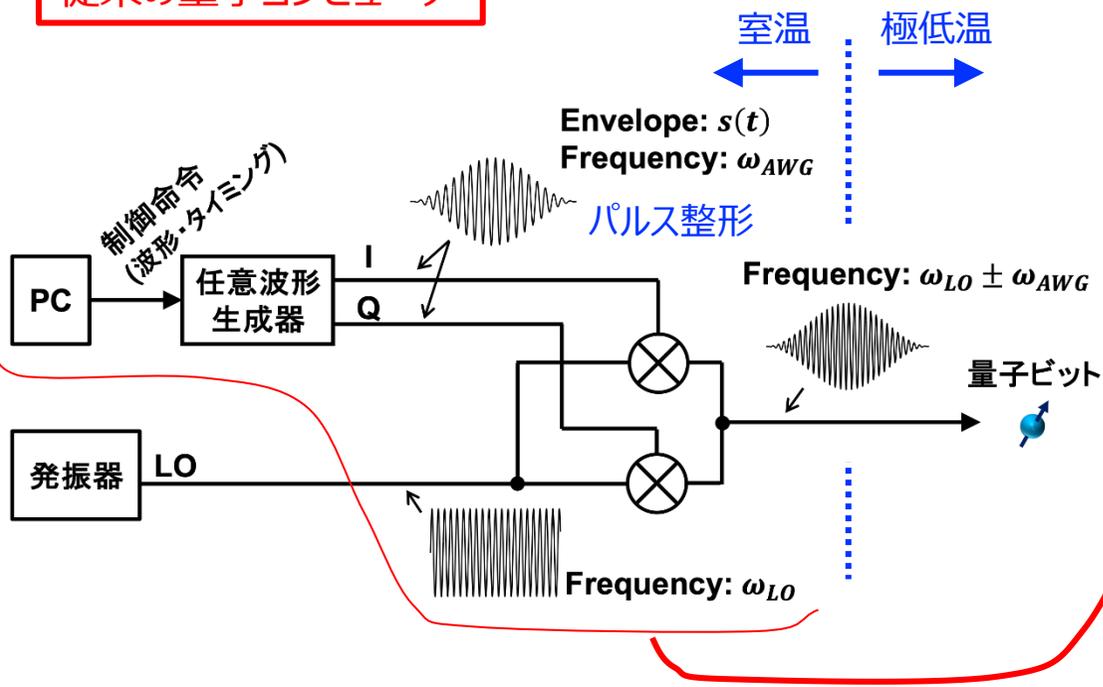
IBMは、超伝導量子コンピュータを世界で初めて商用化するなど、量子コンピュータの研究開発をリードしているが、これまでクライオCMOSの発表はなかった。量子ビット数の拡大に邁進する同社が、それに必要な技術の一つとして、クライオCMOS技術の詳細を公表した事は意義深いと考える。

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM (2/5)

・本発表の概要

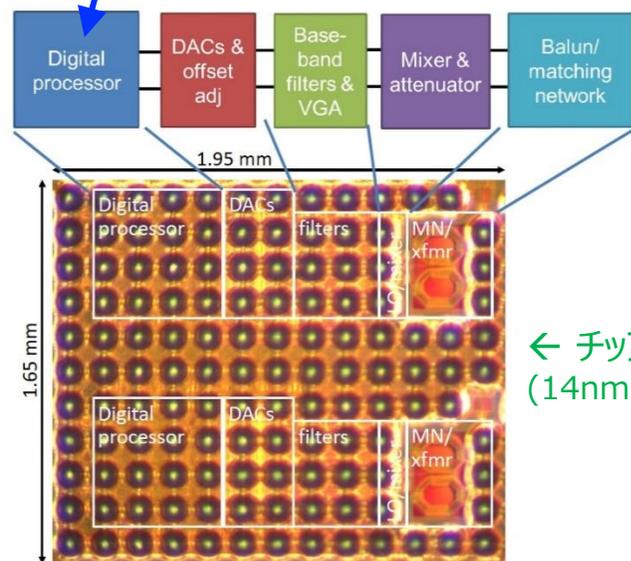
- ・本発表は、超伝導量子ビット制御用CMOS集積回路の提案である。CMOSチップは、14nm FinFETプロセスで製造されている(製造ファブは非公表)。3Kでの動作を実証した。
- ・量子ビットの制御には、パルス整形したマイクロ波(通常、数GHz程度が使用される事が多い。本発表は、4.5-5.5GHz)を量子ビットに印加する。このパルス波形(形状・振幅・パルス幅)や印加タイミングによって、量子ビットをどのように制御するかが決まる。
- ・従来、この量子ビットの制御情報は、室温に置かれたPCなどから送られるが、本提案では、CMOSチップにその機能(プロセッサ)も実装されており、一連の量子ビット制御シーケンスを、冷凍庫内で閉じて実行が可能となる。

従来の量子コンピュータ



提案クライオCMOSチップ

制御用のプロセッサも実装している点が特徴



1チップに集積

← チップ写真 (14nm FinFET)

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM (3/5)

• 技術的に特筆すべき点

- 量子ビット制御用のプロセッサを搭載。従来この部分は室温のPCなどで実行されていたが、提案チップを用いることで、冷凍庫内で完結した量子ビット制御への道が開けた。
- 開発したチップを冷凍機内3K領域に配置、超伝導量子ビットの制御可能である事を実証(量子ビット制御の実証まで実施していない発表も少なくない)。
- 従来に比べて、高い機能を有するが、消費電力は小さい

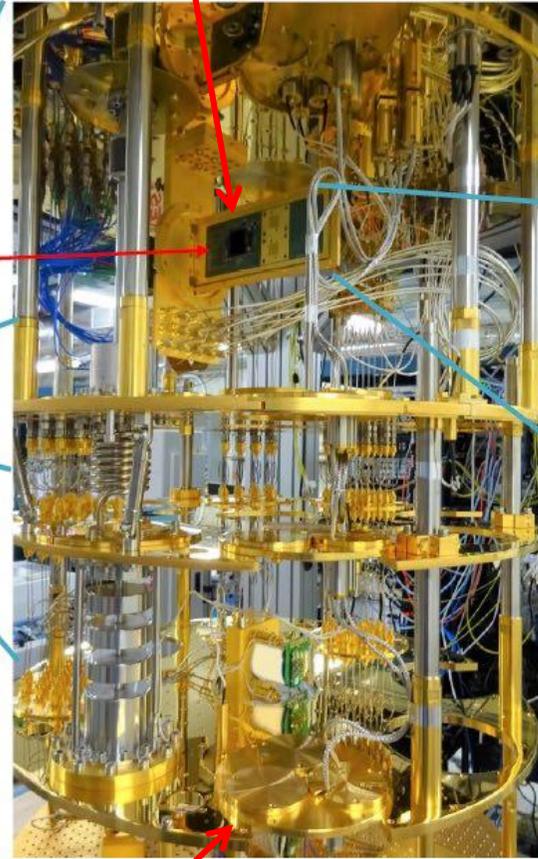
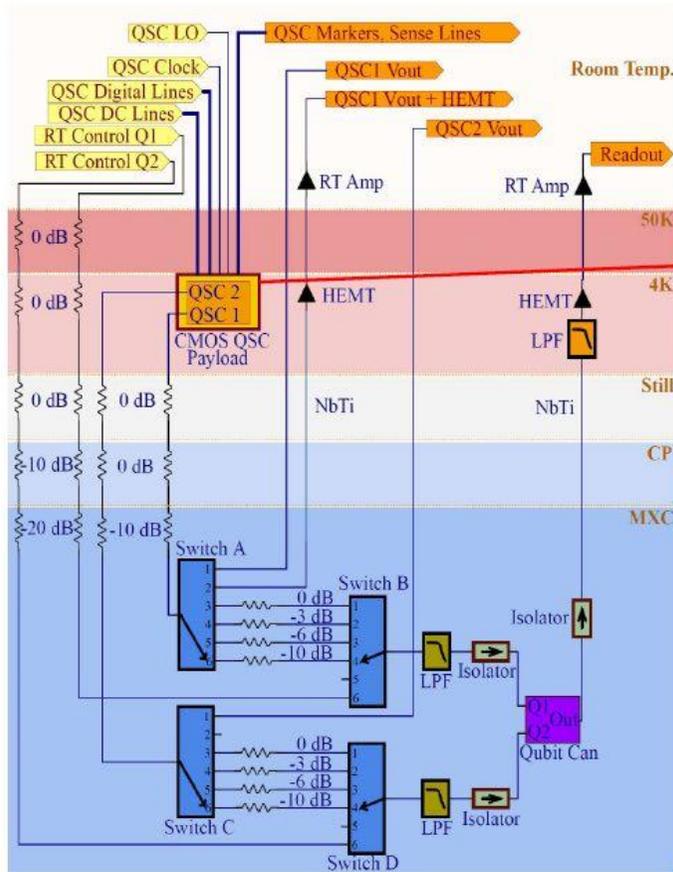
• 実用化までの課題

- 汎用量子コンピュータの実現に向けては、量子計算の「誤り訂正」が必要とされており、その実現には、量子ビットの読み出しの結果を処理しながら制御する必要がある。本発表で提案のクライオCMOSチップは、量子ビットの制御のみ対応しており、読み出しの機能はない。読み出し機能もチップに実装することで、完全に冷凍庫内閉じた制御が実現できる。この点は、今後の課題である。
- また、本提案は低電力である点が強調されているが、1量子ビット当たり23mW消費する。現在の冷凍機の性能では、1W程度の消費電力が上限とされており、その場合、50量子ビットしか制御できない。さらなる回路の低消費電力化、及び、冷凍機性能の向上も、今後の課題である。

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM (4/5)

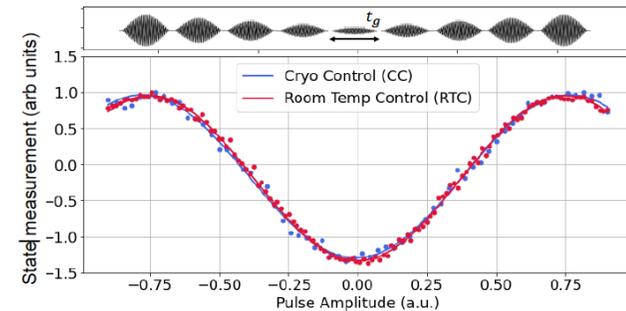
● 測定セットアップ(参考)

提案クライオCMOSチップ



冷凍機

量子ビット制御の測定結果
(ラビ振動を観測)



量子ビット

3-3-1. A Cryo-CMOS Low-Power Semi-Autonomous Qubit State Controller in 14nm FinFET Technology: IBM (5/5)

● 既存研究との比較

	This Work	[4] ISSCC'21	[3] ISSCC'20	[1] ISSCC'19	[2] VLSI-C 2021
Ambient temperature (K)	3	4	3	3	77
Qubit type	transmon	spin	spin & transmon	transmon	transmon
Waveform type	arbitrary	arbitrary	arbitrary	envelope modulated RF pulse	pre-defined and rectangle
RF freq range (GHz)	4.5-5.5	11-17	2-20	4-8	2-7
IF freq range (GHz)	DC-0.3	DC-0.7	DC-0.5	low-freq. envelope	DC-0.6
# channels	2	4	4	1	6
NCOs	updateable phase rotations	64	64	0	1
Sideband method	SSB	SSB	SSB	IQ mod	SSB
DAC speed (GS/s)	1	2.5	1	1	2
DAC bits	10	10	10	~11	9
Output gain/attenuation range	50x	>56x	225x	n/a	n/a
Waveform points	8K	16K	40K	22x16	n/a
Pulse sequence length	4K or unlim	2K	2K	-	n/a
General purpose instruction set	yes, 37 instructions	no (special only)	no (special only)	no	no (special only)
Power/qubit under active control (mW)	23	90 (estimated)	190	2	5.5
Chip area/channel (mm ²)	1.61	~4	4	1.6	0.39
Technology	14nm FinFET	22nm FinFET	22nm FinFET	28nm bulk	40nm bulk

制御用プロセッサを実装しながら、低電力を実現

2-1-10. まとめ

総括

- FinFETの次にくるGAAの発表、FeRAM等の新規メモリの発表、量子コンピュータのためのデバイスの発表、低消費電力でAIを実行するためのIn-memory Computing、3次元実装、異なるChipを組み合わせる一つのチップに実装するChipletによる低消費電力・高性能化など、プロセスの微細化以外の研究開発が盛んに行われていることがわかる。
- Arm、RISC-V、NVIDIAはそれぞれ活用する領域をAIやHPC、さらにはメタバースに広げており、協業する企業のエコシステムを広げている。また、プロセッサ自体のセキュリティに関する発表も多く、ハードウェアのセキュリティが注目されていることがわかる。
- 採択論文数で日中韓をみると、日本は既に韓国に大きく引き離されており、中国にも並びかけられている(VLSI Symposia Circuitでは中国の方が採択論文数が多かった)。発表者所属機関別の発表件数をみると半導体メーカー大手のSamsungからの発表が多く、韓国の研究開発を牽引していることが伺える。

展望

- 今後の研究開発の方向性として、2次元材料トランジスタ、量子コンピューティング、In-memory Computingなどが進められており、さらなる低消費電力・高性能なコンピューティングデバイスが登場すると期待される。
- 中国の採択論文数は増えており、今後学術分野で日本が中国に追い越されることも考えられる。

未来を拓くチカラと技術。



