

## Global Macro Team

허진욱  
Chief Economist  
james\_huh@samsung.com

## Tech Team

황민성  
Team Leader / Senior Analyst  
m.s.hwang@samsung.com

류형근  
Analyst  
hyungkeun.ryu@samsung.com

## Global Equity Strategy Team

박혜란  
Analyst  
hyerani.park@samsung.com

구민정  
Research Associate  
mj.gu@samsung.com

이지혜  
Research Associate  
jhye53.lee@samsung.com

## Compliance Note

본 조사항목은 당사의 저작물로서 모든 저작권은 당사에  
있습니다. 본 조사항목은 당사의 동의 없이 어떠한 경  
우에도 어떠한 형태로든 복제, 배포, 전송, 변경, 대여할  
수 없습니다. 본 조사항목에 수록된 내용은 당사 리서치  
센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나,  
당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따  
라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결  
과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수  
없습니다. 본 자료에는 외부의 부당한 압력이나 간섭 없  
이 애널리스트의 의견이 정확하게 반영되었습니다.

## 글로벌 경기 확장 사이클의 장기화

## 1990s Redux: AI가 주도하는 생산성 개선

- 당사는 1) 통화정책의 침체 대응력 확대, 2) 민간 금융 불균형 부재, 3) 제한적 재정  
긴축 강도, 4) AI 주도의 생산성 개선으로 글로벌 경기 확장이 장기화될 것으로 전망
- 단기적으로는 Tech 섹터의 이익 전망에 긍정적 영향. 한국 메모리 업체 이익 상향  
조정. 삼성전자 목표주가 12만원 상향 및 매수 의견 유지.

## 미국 및 글로벌 경기 확장 사이클이 장기화될 것으로 예상하는 4가지 이유

'20년 하반기부터 시작된 미국 및 글로벌 경기 확장 사이클이 4년째 지속되고 있음. 특히,  
최근 2년 동안 경기 침체에 대한 수많은 우려가 제기되었지만, 아직까지도 미국 및 글로벌  
경제의 침체 조짐은 미미한 상황. 미국 경기 사이클의 장기 추세에서 주목할 변화는 경기  
확장의 기간이 눈에 띄게 길어지는 반면, 경기 수축 기간은 추세적으로 짧아지는 점. 실제  
로 미국 경기 확장 지속 기간은 2차 대전 이전 평균 26개월, 전후~1990년까지 49개월, 그  
리고 1991년 이후 평균 107개월로 두 배씩 길어짐.

당사는 이번 미국 및 글로벌 경기확장 사이클이 장기화될 것으로 전망함. 그 이유는,

첫째, 40년래 가장 빠른 금리 인상 사이클이 경기 침체 없이 일단락되고 금리 인하 전환을  
앞둔 시점에, 높아진 정책금리는 미래의 경기 침체에 대응할 수 있는 중앙은행들의  
firepower가 될 것

둘째, 지난 70년간 Fed 긴축과 oil shock을 제외하고, 미국의 경기 침체를 유발했던 유일한  
요인인 민간 부문의 금융 불균형 징후도 현재 보이지 않음

셋째, 글로벌 금융위기 이후 구조적 저성장을 유발했던 급격한 재정 긴축으로의 전환 가능  
성도 낮음

넷째, 선진국 Tech 투자(AI)가 주도하는 생산성 개선이 골디락스를 장기화하는 요인으로 작  
용할 전망. 퍼낸텔과 통화정책, Tech 사이클 관점에서 당사는 현재의 미국 및 글로벌 경  
제 환경이 90년대 중반, 특히 1995년 초와 상당히 유사한 것으로 판단

## 한국의 메모리 주식과 AI 주도주 투자 권고

반도체의 단기 사이클 (2024~2025년) 전망도 긍정적. 첫째, 기술 방향이 반도체 사용량을  
키워가고 있음. AI 모델은 더욱 빠르고 많은 메모리가 필수. 그렇지 않으면 비싸게 구입한  
프로세서가 놀기 때문. 하드웨어 대역폭의 한계는 교체 수요를 자극 전망. 둘째, 공급 환경  
은 척박한 상황. AI 반도체를 제조하기 위해 낮은 수율과 커지는 사이즈가 문제.

삼성전자와 SK하이닉스의 이익 전망치 상향 조정. 기대감이 최저인 삼성전자의 목표 주가  
를 11만원에서 12만원으로 상향. '좋은 회사'의 조건(높은 마진과 현금흐름, 유지 가능한 경  
쟁력)을 구비한 회사를 투자 필요. 인프라의 엔비디아, 클라우드와 서비스의 마이크로소프  
트와 테슬라가 기술과 시장을 주도하고 있다는 판단.

Risk도 상존. 단기적으로 2000년 IT 버블 붕괴와 같은 상황을 예상하지 않으나, 문제는 AI  
서비스가 소비자의 선택을 받지 못하는 것. 동시에 AI를 기점으로 중국발 첨단 제품의 공급  
과잉 시작 가능성. 단기적으로 소비를 부양하는 정책이 긍정적이나, 기존 공급사에게 위협  
이 될 수 있다는 판단.

## REPORT CONTENTS

---

### 01 글로벌 경기 확장 장기화의 4가지 요인 03p

글로벌 경기 확장 사이클의 장기화: 1990s Redux  
'90년대 후반 미국 골디락스 경험과 주식시장 함의

### 02 Tech: 반도체 지금 더 사세요 25p

AI, Big thing is coming  
삼성전자에 대한 Risk Return이 좋아졌다는 판단의 배경  
Sizing AI. 인공지능 혁신은 상상 이상일 것  
단기 Risk: 2000년 IT 버블 vs. 현재  
장기 Risk: AI를 기점으로 중국발 공급 과잉 시작

### 03 부록 64p

인공지능 기술 개요  
인공지능 밸류체인



## I. 글로벌 경기 확장 장기화의 4가지 요인

### 수많은 우려에도 불구하고, 미국 및 글로벌 경기 확장 사이클이 4년째 지속 중

'20년 하반기부터 시작된 미국 및 글로벌 경기 확장 사이클이 4년째 지속되고 있다. 특히, 최근 2년 동안 경기 침체에 대한 수많은 우려가 제기되었지만, 아직까지도 미국 및 글로벌 경제의 침체 조짐은 미미한 상황이다.

미국 경기 사이클의 장기 추세에서 주목할 변화는 경기 확장 기간이 눈에 띄게 길어지는 반면, 경기 수축 기간은 추세적으로 짧아지고 있다는 점이다. 실제로 미국 경기 확장 지속 기간은 2차 대전 이전 평균 26개월, 전후~1990년까지 49개월, 그리고 1991년 이후 평균 107개월로 두 배씩 길어지고 있다. 이는 미국 경기 사이클에서 normal한 상황은 경기 확장 국면이고, 그 길이는 지속적으로 길어지는 반면, 경기 침체는 상당히 예외적인 기간에 해당한다는 점을 잘 보여준다.

### 미국 및 글로벌 경기 확장 사이클의 장기화를 예상하는 4가지 이유

당사는 현재 진행 중인 미국 및 글로벌 경기 확장 사이클이 장기화될 가능성이 높은 것으로 판단한다. 그 이유는 첫째, 40년래 가장 빠른 금리 인상 사이클이 경기 침체 없이 일단락되고 금리 인하 전환을 앞둔 시점에, 높아진 정책금리는 미래의 경기 침체에 대응할 수 있는 중앙은행들의 firepower가 될 것이다. 둘째, 지난 70년간 Fed 긴축과 oil shock을 제외하고, 미국의 경기 침체를 유발했던 유일한 요인인 민간 부문의 금융 불균형 징후도 현재 보이지 않기 때문이다.

셋째, 글로벌 금융위기 이후 구조적 저성장을 유발했던 급격한 재정 긴축으로의 전환 가능성도 낮다. 넷째, 선진국 Tech 투자(AI)가 주도하는 생산성 개선이 골디락스를 장기화시키는 요인으로 작용할 전망이다. 펀더멘털과 통화정책, Tech 사이클 관점에서 당사는 현재의 미국 및 글로벌 경제 환경이 90년대 중반, 특히 1995년 초와 상당히 유사한 것으로 판단한다.

### 주식시장 관점에서 현재의 AI 붐은 과거 90년대 후반 IT 버블과는 차별화됨

주식시장 참여자들이 궁금해하는 건 크게 2가지이다. 첫째, AI 붐이 90년대 후반 IT 버블 경로를 따를 것인가 하는 것과 둘째, 90년대 후반이 아닌 중반의 재현이라면 시장과 AI 주식을 어떻게 접근해야 할 것인가이다. 버블 붕괴는 기대와 현실이 어긋나는 데에서부터 시작한다. AI 주식들은 실제로 기대를 넘어 증명하고 있다는 점이 IT 버블과의 가장 중요하고도 큰 차이이다. 따라서 향후 실망이 발생하더라도 IT 버블과 같은 전철을 밟을 위험은 적다.

다만, AI를 중심으로 한 시장 랠리가 지속되면서 업종 쏠림이 30년래 최고 수준에 도달한 점은 우려된다. 이러한 쏠림 현상은 결국 균형을 찾기 마련인데, 주요 catalyst는 금리 인하에 대한 기대 회복과 미국의 장기 골디락스 환경이 될 것이다. 그간 가격이 계속 부진했고, 향후 실적 개선세가 긍정적인(vs. AI 실적 모멘텀은 상대적으로 둔화 예상) 업종/주식으로 점차 시선을 넓혀가는 것이 적절한 투자 전략으로 판단한다.

## 글로벌 경기 확장 사이클의 장기화: 1990s Redux

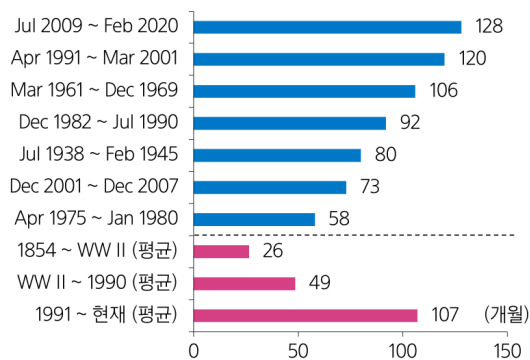
## 글로벌 경기 확장 사이클의 장기화를 예상하는 4가지 이유

비경제적(팬데믹) 요인에 따른 전례 없는 경기 침체와 주기적인 lockdown 반복 등으로 팬데믹 초기, 이번 경기 침체 충격에서 벗어나는 회복 사이클은 매우 느리고 더디게 진행될 것이라는 전망이 주를 이루었다. 그러나, 글로벌 금융위기(GFC) 때와는 달리, 재정정책이 주도하고 통화정책이 보조했던 정책 대응은 경제 정상화 과정에서 전례 없는 회복력을 보여주었다. 이에 따라, 주요국에서 팬데믹의 주요 충격이 GDP 대비 최대 약 10~15%로 막대했지만, 초기 우려와는 달리 팬데믹 이전의 GDP를 넘어서는 데 미국과 중국은 불과 3~4분기, 유로존도 7분기밖에 걸리지 않았다.

2020년 하반기부터 시작된 글로벌 경기 회복 사이클은 '21년 하반기 이후 발생한 인플레이션 급등과 이에 대응한 주요국들의 동반 금리 인상으로 이어졌다. 당사는 미국 및 글로벌 경제가 지난 4년간의 정상화 과정을 거쳐, 올 하반기부터는 정상 국면에 진입(from normalization to normal)할 것으로 예상된다. 비록 물가 목표 2%로 완전히 복귀한 것은 아니지만, 팬데믹 이후의 전례 없는 공급 측 제약(제조업 공급망 차질, 유럽 에너지 가격 급등, 주요국 노동공급 제약 등)에 따른 인플레이션 급등 국면이 상당 부분 마무리되었다. 이에 따라, 미국과 유로존의 핵심 인플레이션은 이제 3%를 하회하기 시작하였고, 주요 선진국들의 올해 말/내년 초 핵심 인플레이션 상승률은 물가 목표에 상당히 근접하며 정상화될 것으로 예상된다. 이는 글로벌 경제가 성장률, 물가, 노동시장 등 대부분의 분야에서 잠재 수준에 근접하는, 보다 지속 가능한 정상적인 경기 확장 국면으로 복귀할 것임을 의미한다.

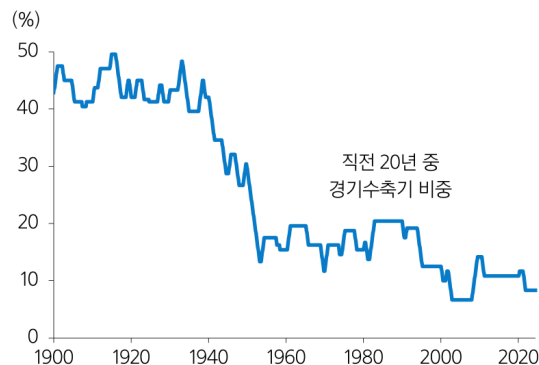
이에 따라, 40년래 가장 빠른 속도로 진행된 주요국 금리 인상 사이클이 대부분 지난해 3Q 중 마무리 되었고, 오는 6월부터 ECB와 BOE를 시작으로 9월에는 Fed까지 가세하면서, 중립금리 수준을 향한 점진적인 금리 정상화(인하) 사이클이 시작될 전망이다. 지난 2년간 미국을 중심으로 경기 침체에 대한 수많은 우려가 끊임없이 제기되었던 주된 원인이 급격한 금리 인상이었다는 점에서, 오는 6월부터 당사 및 시장의 예상대로 주요국의 점진적인 금리 인하가 진행될수록, 경기 침체 가능성은 비례해서 낮아지는 반면, 골디락스에 대한 기대는 지속적으로 강화될 것이다. 이것이 당사가 미국 및 글로벌 경기 확장 사이클의 장기화를 예상하는 첫 번째 핵심 이유이다. 다음으로 지난 70년간 Fed 통화긴축과 oil shock을 제외하면, 미국의 경기 침체를 유발했던 거의 유일한 요인인 민간 부문(가계 및 기업)의 금융 불균형 징후도 현재 전혀 보이지 않는다. 세 번째로 주요국 모두에서, 글로벌 금융위기 이후 구조적 저성장(secular stagnation)을 유발했던 급격한 재정 긴축으로의 전환이 재현될 가능성도 높지 않다.

## 추세적으로 길어지고 있는 미국의 경기 확장 사이클



자료: NBER, 삼성증권

## 반면, 경기 둔화 및 침체기는 추세적으로 짧아짐



자료: NBER, 삼성증권

끝으로, 보다 중장기적인 관점에서 당사는 현재의 미국 및 글로벌 경제 환경이 1990년대 중반, 특히 1995년 초와 상당히 유사한 것으로 판단한다. '91년 초부터 '21년 초까지 이어진 정확히 10년간의 미국 경기 확장 사이클은 크게 두 국면으로 나눌 수 있다. 1980년대 후반 누적된 Fed의 금리 인상과 '90년 하반기 걸프전 발발에 따른 유가 급등 등이 겹쳐 짧은 경기 침체를 겪은 이후, 미국 경제가 정상화되고 노동시장이 완전고용에 진입하면서, '94년 2월~'95년 2월까지 1년간 300bp의 급격한 금리 인상이 진행되었다. 이는 팬데믹으로 인한 짧은 경기 침체 이후, '22년 초부터 '23년 3Q까지 진행된, Fed가 주도했던 글로벌 통화긴축 사이클과 매우 유사하다. 이후 1990년대 초반부터 누적된 Tech 투자가 '95년 이후 본격적인 생산성 개선으로 이어지면서, '95~'00년까지 미국의 골디락스(경기 확장+안정된 인플레이션) 국면이 지속되며, 10년간의 장기 경기 확장 사이클이 진행된 것이다.

1991년 2Q부터 정확히 10년간 이어진 1990년대 장기 확장 사이클에서 핵심은 Tech 투자의 확대를 통한 생산성 향상을 통해 미국의 총공급 능력, 즉 잠재 성장률이 상승함에 따라, 인플레이션의 상승을 유발하지 않고 더 높은 성장을 달성할 수 있었다는 점이다. 이러한 관점에서 지난 3월 FOMC의 미국 경제전망 수정과 파월 의장의 기자회견 내용은 팬데믹 이후 미국 경제의 공급 능력 확대 가능성을 처음으로 인정했다는 점에서 매우 중요한 의미를 갖는다.

한편, 미국 경기 사이클의 장기 추세에서 주목할 만한 변화는 경기 확장 사이클의 지속 기간이 시간이 갈수록 눈에 띄게 길어지는 반면, 경기 수축기가 차지하는 비중은 추세적으로 낮아지고 있다는 점이다. 실제로 미국의 경기 확장 지속 기간은 2차 대전 이전에 평균 26개월, 전후~1990년까지 49개월, 그리고 1991년 이후 평균 107개월로 두 배씩 추세적으로 길어지고 있다. 반면, 경기 둔화/침체 기간은 2차 대전 이후 추세적으로 짧아지면서, 평균 10개월에 불과하다. 이는 통화/재정정책 수단의 정교화, 기업들의 Just-In-Time 재고 관리도입, 경제지표의 적시성과 정확성 개선, 금융 감독 기능 강화 등 다양한 요인에 근거한다. 이는 미국 경기 사이클에서 normal한 상황은 경기 확장 국면이고, 그 길이는 지속적으로 길어지는 반면, 경기 침체는 상당히 예외적인 기간에 해당한다는 점을 잘 보여준다.

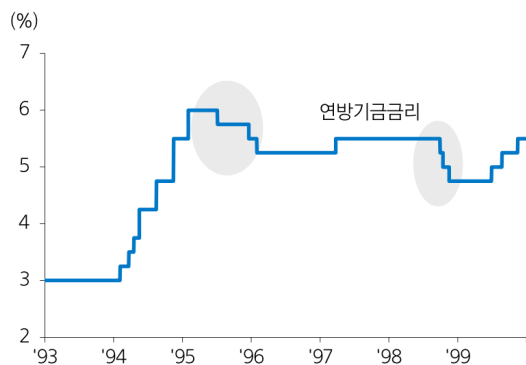
### 첫째, 미래의 경기 침체에 대응할 수 있는 통화정책의 충분한 firepower (From headwind to tailwind)

미국 및 글로벌 경기 확장 사이클의 장기화를 예상하는 가장 중요한 첫 번째 요인은 통화정책이다. '22년 상반기부터 시작된 주요국들의 급격한 금리 인상은 다행스럽게도 아직까지 우려했던 경기 침체를 유발하지 않았다. 따라서, 주요국 중앙은행들의 금리 인상 사이클이 마무리되고 점진적인 금리 인하를 앞둔 현시점에, 높아진 정책금리는 이제 미래의 경기 침체에 대응할 수 있는 중앙은행들의 firepower가 될 것이다. 다시 말해, 주요 중앙은행들이 물가 정상화에 대한 확신이 높아져 금리 인하를 개시한 이후부터는, 경기 하방리스크에 보다 적극적으로 대응할 것이다.

특히, 지난 3월 FOMC 이후 Fed가 조기 금리 인하를 할 수 있는 조건으로 언급하기 시작한 “노동시장의 예상치 못한 악화”가 발생할 경우, 적어도 보험적 성격의 insurance cuts을 제공할 가능성이 높다. 5월 FOMC 기자회견에서 파월 의장은 비록 최근 둔화 속도가 예상에 미치지 못하는 못하지만, headline과 핵심 PCE 물가가 모두 3%를 하회하기 시작하면서, Fed는 또 다른 정책 목표인 ‘완전 고용’에도 초점을 맞추기 시작했다고 언급하였다. 이것은 지금까지 ‘물가 안정’만을 최우선 순위로 두고 정책을 결정해 온 것과는 달리, 향후에는 고용 지표의 추가 악화 여부에도 신경을 쓰겠다는 의미이다. 이는 1994~95년 초까지의 급격한 금리 인상을 끝낸 Fed가 1995년과 1998년 두 차례의 insurance cuts을 제공했던 것을 떠올리게 한다. Fed는 반드시 경기 침체를 우려하지 않는 상황에서도 대외 여건의 악화나 금융시장의 불안이 확산될 경우, 혹시 있을 수 있는 경기 침체 리스크를 선제적으로 완화하기 위한 보험적 성격의 금리 인하를 제공해 왔다. 1995년 멕시코 위기, 1998년 LTCM 사태, 그리고 2019년 중국/유로존 경기 악화 등이 대표적인 사례이다.

비단 Fed뿐 아니라, ECB나 BOE 등 대부분의 주요국 중앙은행 정책금리가 지난 2000년 이후 가장 높은 수준에 있다는 점은 향후 글로벌 경기 침체에 대응한 통화정책 공조를 가능케하고, 이는 2020년 하반기부터 시작된 미국 및 글로벌 경기의 확장 사이클을 장기간 유지시킬 수 있는 핵심 요인이 될 것이다. 특히, 팬데믹 이후 경제 정상화 과정에서 발생한 인플레이션 급등에 대응하기 위해 주요 중앙은행들의 동반 금리 인상이 급격히 진행되면서 지난 2년간 미국 및 글로벌 경기 침체 우려가 정점에 달했다. 그러나, 주요국 중앙은행들의 금리 인상 사이클은 이미 지난해 3Q 중 일단락되고, 올 6월부터 ECB와 BOE를 시작으로 중립금리를 향한 점진적인 금리 정상화(인하) 사이클이 시작될 가능성이 매우 높은 상황이다.

### 급격한 긴축 이후 경기 침체를 방지하기 위한 insurance cuts 사례



참고: 음영부분은 보험적 성격의 insurance cuts  
자료: Federal Reserve, 삼성증권

### 다음 경기 침체 시, Fed는 QE 없이 금리 인하만으로 대응 여력 충분

경기 침체 시기	금리 인하 폭 (% 포인트)	핵심 PCE 물가 (%)	실업률 갭 (% 포인트)
57년 8월~58년 4월	2.9	3.2	-1.3
60년 4월~61년 2월	2.8	2.1	-0.3
69년 12월~70년 11월	5.5	4.8	-2.4
73년 11월~75년 3월	7.7	10.2	-1.4
80년 1월~80년 7월	4.8	9.1	-0.2
81년 7월~82년 11월	10.4	8.8	1.0
90년 7월~91년 3월	5.3	4.3	-0.4
01년 3월~01년 11월	4.8	2.0	-0.8
07년 12월~09년 6월	5.1	2.3	0.1
20년 2월~20년 4월	1.5	1.8	-0.4
역사적 평균	5.1	4.9	-0.6
현재	5.4	2.8	-0.1

자료: Federal Reserve, 삼성증권

Fed 또한 비록 금리 인하 시점이 당초 기대보다 다소 지연되기는 했지만, 5월 FOMC를 통해 파월 의장은 “현재의 정책금리는 충분히 긴축적인 수준이며, 정책금리의 다음 방향성이 인상일 가능성은 낮다”는 점을 분명히 했다. 이는 사실상 금리 인상의 threshold가 매우 높다는 점을 확인시켜준다. 파월 의장은 현재의 금리 수준이 충분히 긴축적이지 않아 금리 인상이 필요하다고 판단할 수 있는 조건에 대해서는 “totality of all the things”를 강조하였다. 즉, 단순히 물가 지표뿐만 아니라, 기대 인플레이션과 고용 지표 등 여타 다양한 지표들을 종합적으로 고려해서 판단해야 한다는 뜻이다. 반면, 금리 인하에 대해서는 “분명한 것은 향후 금리 인하를 결정하는 데 있어서 그 핵심에 있는 것은 물가 지표이다(Clearly, incoming inflation data would be at the very heart of that decision)”라고 사실상 물가 지표만을 강조하였다. 결국, 파월 의장이 언급한 정책금리 변화의 threshold 높이를 비교하자면, 금리 인상 >> 금리 인하 > 금리 동결로 볼 수 있다.

현재의 미국 매크로 환경과 통화정책 스탠스는 불과 12개월 만(1994년 2월~1995년 2월)에 3.0%에서 6.0%까지 300bp에 달하는 급격한 금리 인상 사이클이 끝난 1995년 초 상황과 매우 유사하다. 지난 1994~95년의 금리 인상 사이클의 주된 배경은 다음과 같다. 미국 저축대부조합(S&L)에 대한 구조조정이 상당 부분 일단락되고, 1차 걸프전이 단기간에 종료됨에 따라, 1992년 들어 미국의 경제 성장률과 실업률이 Fed의 예상을 크게 상회하는 급격한 개선 추세를 지속하였다. 당시 미국의 인플레이션율은 2%대로 현재와는 달리 매우 안정된 수준을 유지하고 있었다. 그러나, Fed는 4%대의 경제 성장률과 지속적인 가동률 상승, 특히 실업률이 6.5%를 하회하기 시작함에 따라, 경제 내 slack 축소로 인플레이션 압력이 빠르게 상승할 것으로 판단하였다. 이에 따라, 1994년 2월 3.0%에서 12개월 동안 6.0%까지 무려 300bp에 달하는 금리 인상을 단행하게 된다.

그러나, 1995년 초 이후 인플레이션이 2% 내외에서 안정세를 유지하는 가운데, 멕시코 위기 등 해외 여건이 악화되면서 미국의 성장률이 1%중반 수준으로 둔화되자, 혹시 모를 경기 침체 리스크를 사전에 완화시키기 위해 '95년 7월 부터 '96년 1월까지 3차례에 걸쳐 75bp의 insurance cuts을 단행한다. 이후 미국 경기하방 리스크가 완화되자 '98년까지 장기간 금리 동결 기조('97년 3월 25bp 소폭 인상)를 유지했던 Fed는 아시아 외환위기에 이은 LTCM 사태 발생에 대응하여 '98년 9월부터 11월까지 다시 한번 3차례에 걸쳐 75bp의 insurance cuts을 시행한다.

한편, 미국의 경우, 한국전쟁 이후 10번의 경기 침체기의 평균 정책금리 인하 폭이 약 5.1%포인트로 공교롭게도 현재의 정책금리 수준 5.4%와 거의 유사하다. 그런데, 최근 10번의 경기 침체에서 현재의 금리 인하 여력인 5.4%포인트보다 더 큰 폭의 금리 인하가 필요했던 사례는 지난 1973년과 1981년의 1, 2차 오일 쇼크밖에 없다. 이는 현재 배럴당 80달러 내외인 국제유가 수준에서 3차 오일 쇼크가 발생할 가능성이 매우 낮음을 감안할 때, 언제 올지는 모르지만 다음 번 경기 침체 때 비전통적 수단인 QE를 사용하지 않고, 금리 인하만으로도 충분한 정책 대응이 가능할 것임을 시사한다. 그 확률은 80%에 달한다.

이는 현시점에서 매우 중요한 시사점을 제공한다. 지난해 7월을 마지막으로 40년래 가장 빠른 속도의 Fed 금리 인상이 일단락되고, 1년 가까이 금리 동결 기조가 유지되고 있다. 앞서 언급한 대로, 최근의 인플레이션 지표 반등에도 불구하고, FOMC의 추가 인상의 허들은 매우 높다는 것이 확인된 상황이다. 또한, 파월 의장은 5월 기자회견에서 이제 ‘물가 안정’ 일변도에서 또 다른 정책 목표인 ‘완전 고용’에도 신경을 쓰기 시작했음을 밝혔다. 이는 향후 1) 실업률이 4% 이상으로 추세적으로 상승하거나, 2) 신규 취업자수가 빠르게 둔화 혹은 감소할 경우, Fed가 큰 폭의 금리 인하는 아니더라도 최소한 지난 '95년이나 '98년과 같이 insurance cuts을 제공할 가능성이 높아지고 있음을 시사한다. 이것은 미래의 경기 침체에 대응할 수 있는 주요국 중앙은행들의 높아진 금리 인하 여력(firepower)을 감안할 때, 시간이 갈수록 미국 및 글로벌 경제의 침체 리스크가 낮아질 것임을 의미한다.

### 둘째, 경기 침체를 유발할 수 있는 민간 부문의 금융 불균형 징후 미미

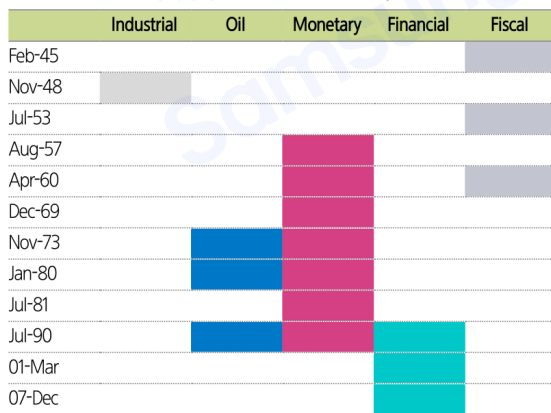
*"None of the post-war expansion died of old age. They were all murdered by the Fed (2차 대전 이후의 미국 경기 확장은 기간이 길어져서 끝난 것이 아니다. 모두 Fed의 긴축에 의해 살해됐다)"*라는 MIT Dornbusch 교수의 유명한 말처럼, 경기 확장 사이클은 기간의 문제가 아니라, Fed의 긴축이나 민간 부문의 금융 불균형, 오일 쇼크와 민간 부문의 금융 불균형이 각각 20%씩(복수 요인 허용)을 차지한다.

특히, 지난 70년간 미국의 경기 침체를 유발했던 핵심 요인을 살펴보면, 약 70%가 Fed의 통화 긴축이고, 그 외에는 오일 쇼크와 민간 부문의 금융 불균형이 각각 20%씩(복수 요인 허용)을 차지한다.

이 중 민간 부문의 금융 불균형은 가계나 기업의 금융수지가 적자인지, 아니면 흑자인지를 말한다. 쉽게 말해, 한 국가의 가계나 기업이 벌어들이는 내부 현금으로 현재의 소비나 투자를 감당할 수 있는가를 의미한다. 일반적으로 가계 저축이 기업 투자보다 많으면 민간 금융수지가 흑자를, 적으면 민간 금융수지가 적자를 기록하게 된다. 민간 금융수지는 국민소득 항등식에 의해, 경상수지에서 재정수지를 차감하여 구할 수 있다. 민간 금융수지 적자가 심화될수록, 민간 부문의 금융 불균형이 악화되면서 경기 침체를 유발하거나, 혹은 경기 침체가 깊고 길어지는 원인이 되는 것이다.

당사가 향후 미국의 경기 침체 가능성을 점검하는 데 있어서, 민간 부문의 금융 불균형에 특히 주목하는 이유는 이것이 팬데믹을 제외하면, 가장 최근 경험했던 두 번의 미국 및 글로벌 경기 침체와 금융위기를 유발했던 핵심 요인이기 때문이다. 바로 '00~02년의 Dot-com 버블과 '07~08년 Sub-prime 사태이다. 이미 널리 알려진 대로, Dot-com 버블 때는 기업들의 과도한 투자 확대로 민간 금융수지 적자가 GDP 대비 -6%를 상회하는 수준까지 악화되었고, Sub-prime 위기 때는 가계의 과도한 부채 부담으로 금융수지 적자가 -4%를 상회하였다. 이러한 민간 부문의 금융 불균형 악화는 결국 미국을 넘어 깊은 글로벌 경기 침체와 금융위기로 이어졌다.

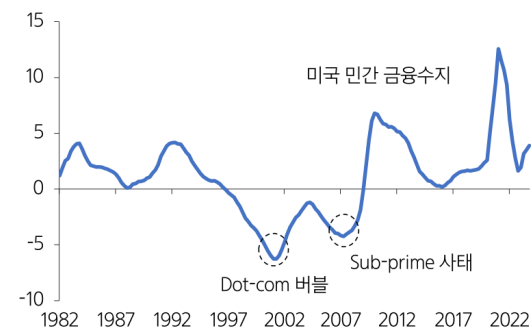
### 2차 세계 대전 이후 美 경기 침체의 원인: Fed, Oil and 금융 불균형



자료: 삼성증권

### 미국 민간 부문의 과도한 금융 불균형 조짐 없어

(GDP대비, %)



자료: CEIC, 삼성증권

팬데믹 직전 GDP 대비 평균 +1~2%의 흑자를 유지하던 미국의 민간 금융수지는 팬데믹 초기의 막대한 재정 및 통화정책의 지원에 힘입어, '21년 중 흑자 규모가 GDP 대비 +10%를 상회하는 수준까지 급증했다. 그러나 이후 지원 프로그램의 종료, Fed의 급격한 금리 인상과 이에 따른 금융 여건의 긴축, 인플레이션 급등 등으로 금융수지 흑자가 지난해 말 기준 GDP 대비 +3.9%까지 둔화된 상황이다. 그러나, 보다 정확히 표현하면 정부 지원에 힘입어 일시적으로 급등했던 미국 민간 금융수지가 보다 지속 가능한 수준으로 정상화된 것이다. 이는 과거 Dot-com 버블이나 Sub-prime 사태 직전의 적자와는 비교할 수 없을 만큼 양호한 수준이며, 심지어는 장기 평균(+3.6%)을 상회하며 개선 추세를 시현하고 있다. 이는 가계 부문의 경우, 노동시장 호황에 힘입어 견고한 임금 및 소득증가세가 지속되고 있기 때문이며, 기업의 경우도 글로벌 경기 확장의 지속으로 양호한 이익 개선 및 재무건전성을 유지하고 있기 때문이다.

비록 올들어 Fed의 금리 인하 시점이 당초 기대보다 지연되고 있지만, 현재 미국의 금융 여건 지수는 여전히 장기 평균 수준을 기록하고 있다. 이는 민간 부문의 전반적인 자금조달 및 신용 여건이 너무 긴축적이지도, 그렇다고 너무 완화적이지도 않음을 의미한다. 따라서, 향후 Fed가 예상과 달리 큰 폭의 추가 금리 인상을 단행하지 않는다면, 금리 인하 개시가 다소 지연되더라도 미국 민간 부문의 전반적인 금융 여건과 견고한 금융수지 흑자 기초가 급격히 악화될 가능성은 매우 제한적인 것으로 평가한다.

한편, 실질유가를 감안할 때, 당사는 70년대 1~2차 오일쇼크와 같은 stagflation을 유발할 수 있는 국제유가 수준을 최소 120달러 이상으로 추정한다. 따라서, 최근 러시아-우크라이나 전쟁과 중동 지역의 지정학 리스크 확대에도 불구하고, 여전히 WTI 기준 배럴당 80달러 내외를 유지하고 있는 현재의 유가 수준에서 미국 및 글로벌 경기 침체를 유발할 수 있는 3차 오일 쇼크가 발생할 확률은 매우 낮다. 결론적으로, Fed의 통화긴축을 제외하고, 지난 70년간 미국의 경기 침체를 유발했던 나머지 두 리스크인 민간 금융 불균형과 유가 급등 모두, 현시점에서는 거의 관찰되지 않는다. 이것이 당사가 미국 및 글로벌 경기 확장 사이클의 장기화를 예상하는 두 번째 이유이다.



## 셋째, 글로벌 금융위기(GFC) 직후와는 달리, 주요국 모두 급격한 재정 긴축 전환 가능성 낮아

팬데믹 이전부터 진행되던 인구 고령화와 베이비 붐 세대의 은퇴가 팬데믹을 거치면서 주요 선진국 모두에서 더욱 가속화됨에 따라, 경제활동 인구의 감소를 상쇄할 수 있는 생산성 제고가 가장 큰 당면 과제로 떠오르고 있다. IMF도 Fiscal Monitor 최신판에서 “전 세계적으로 생산성과 경제 성장을 제고할 수 있는 기술의 혁신과 보급(diffusion)을 활성화시킬 수 있도록 정교하게 설계된 재정정책이 요구된다”고 강조하고 있다. 이것은 약 15년 전인 글로벌 금융위기(GFC) 직후 강력한 재정 긴축을 권고했던 것과는 사뭇 달라진 입장이다. 그 배경은 당시 IMF의 재정 긴축 처방으로 인해 실업이 장기화되는 이력효과(hysteresis)가 발생하여, 오히려 경기 회복을 크게 지연시켰다는 비판이 거세졌기 때문이다. 실제로 미국과 유로존은 장기간 1% 내외의 저성장이 이어지며, GFC 이전의 실업률로 복귀하는 데 각각 10년과 12년이나 소요되었다.

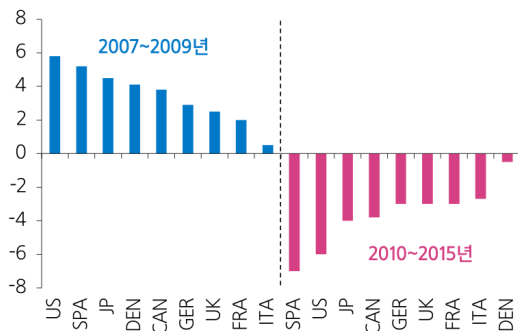
2008~9년 글로벌 금융위기(GFC)를 해결하는 과정에서 확장적 통화-재정정책의 공조가 진행되면서 금융기관들이 보유하고 있던 민간 부채 부담이 정부 부채로 이전되었다. 그러나 2010년부터 주요국 경기가 빠르게 경기 정상화에 진입하면서, 주요 선진국에서 일시적으로 확장 전환했던 재정정책 기조가 빠르게 긴축 혹은 균형 재정으로 전환되었다. 특히, 2011~2013년 유럽의 재정위기와 2012~14년 미국 오바마 정부의 sequester와 재정 절벽으로 대표되는 재정 긴축을 거치면서, 경기 부양을 통화정책에만 의존하는 Monetary dominance의 시대가 2019년까지 지속되어 왔다.

당사는 이러한 통화정책과 재정정책의 엇박자가 GFC 이후 10년간 이어진 만성적 총수요 부족과 잠재성장률의 하락(Secular stagnation: 구조적 저성장)의 주된 원인으로 판단한다. 이에 따라, 글로벌 교역이 둔화되는 가운데, 경기 부양을 위해 각국이 통화 완화 기조를 강화할수록 경쟁적 통화 절하(환율 전쟁) 논란이 격화되고, 무역 갈등이 심화되었다. 이는 Monetary dominance가 지니는 축소 지향적이고 갈등적인 특성에 기인한다.

그러나, 팬데믹 충격은 균형 재정에 얽매어 장기간 잊힌 재정정책의 필요성과 중요성을 각성시키는 계기가 되었다. 코로나로 인한 수요 충격 앞에서 재정정책 사용의 가장 큰 걸림돌인 ‘도덕적 해이(Moral hazard)’ 이슈를 누구도 제기하지 않았다. 팬데믹 이후 미국과 유럽 등 주요국들은 재정 확대를 통한 유효수요 창출과 잠재 성장을 제고에 주력하고 있다. 이미 주요국 정부는 IT, 친환경, 의료 등 인프라 투자에 집중하며, 민간 부문의 투자 확대를 유도하기 시작하였다. IMF는 GDP 1%의 공공 투자 확대가 향후 2년간 최대 약 2.7%의 GDP, 10%의 민간 투자, 1.2%의 고용 증가를 유발할 것으로 추정하고 있다.

## GFC 이후의 급격한 재정 긴축이 Secular stagnation을 유발

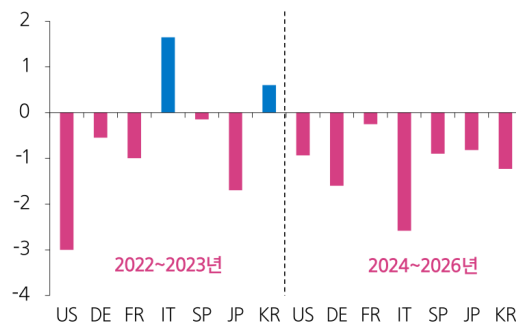
(GDP대비, %)



참고: 재정 자극이 (-)일수록 긴축 재정 기조  
자료: IMF, 삼성증권

## 주요국 대부분 향후 3년간 최근의 재정 긴축 강도를 유지할 전망

(GDP대비 평균 재정자극, %)



참고: 재정 자극이 (+)일수록 확대 재정 기조  
자료: IMF, 삼성증권



미국은 러시아-우크라이나 전쟁과 이스라엘-하마스 전쟁 등으로 예상보다 국방 관련 재정지출이 크게 증가한 상황이다. 반면, 저소득층에 대한 Medicaid 지출은 지난해보다 축소됨에 따라, 국방 지출 확대의 효과를 상당 부분 상쇄하고 있다. FY23년에 비해 FY24년 들어 학자금 지원이나 기타 지출이 감소하면서, 재정정책의 기조를 보여주는 primary 재정수지(이자지급 제외)의 적자는 GDP 대비 약 2%포인트나 축소가 예상된다. 다만, GDP에 직접적으로 기여하는 재정지출의 축소 정도는 상대적으로 제한적이어서, 올해 예상되는 재정 자급은 소폭의 (-)에 그칠 전망이다. 또한, 국채금리의 평균 수준이 상승하면서, 이자지급까지 포함한 전체 재정적자 규모도 GDP 대비 -6.3%를 기록했던 지난해와 거의 유사한 수준을 유지할 것으로 예상된다.

향후 미국의 재정정책 스탠스는 오는 11월 치러지는 대선과 의회 선거 결과에 따라 다소 유동적이다. 다만, 백악관-상원-하원이 어떤 결과가 나오더라도 상대당의 협조 없이 단독으로 세금 인상 혹은 인하, 지출 확대 혹은 축소를 처리하기는 어렵다. 따라서, 의회 내 합의 과정에서 재정 스탠스는 어느 방향으로든 급격하기보다는 점진적이고 완만한 수준의 변화가 예상된다.

특히, 유럽의 경우, 러시아-우크라이나 전쟁 발발 이후, 러시아에 대한 에너지 의존도 감축과 국방력 강화를 위한 재정지출이 크게 증가한 상황이다. 이는 이전부터 유럽이 주도해 온 지구온난화 등 기후변화에 대응하기 위한 그린 투자와 더불어, 유럽 국가들이 더 이상 균형 재정을 고수하기 어렵게 만드는 변화 요인이 되고 있다. 더욱이, 올해 말 미 대선에서 트럼프가 당선될 경우, NATO에 의존해 온 방위력을 자체적으로 증강시키기 위한 국방지출이 장기간에 걸쳐 추세적으로 확대될 가능성이 높다. 유로존 중심국들 중 프랑스와 이탈리아의 경우, EU 재정규율을 준수하기 어려운 상황이지만, 비록 패널티를 받더라도 경기충격을 감내하면서까지 급격한 긴축 재정을 통한 재정건전화를 추진할 가능성은 제한적인 것으로 예상된다.

한편, 중국 정부는 단기적으로는 4.5~5.0% 수준에서 경기를 안정화시키는 것과 함께 중장기적으로는 주택시장의 공급과잉을 연착륙시키는 것을 최우선 정책 과제로 삼고 있다. 특히, 팬데믹 이후 내수 부진이 심화되면서 디플레이션 압력이 높아지는 상황에서, 중국 정부는 상당 기간 일정 수준의 재정적자를 용인할 수밖에 없는 상황이다. 지난 3월 양회에서 올해 GDP 대비 재정적자 목표 비율은 지난해와 동일하게 -3.0%로 제시되었고, 지방정부 특수채 발행 3.9조 위안으로 지난해(3.8조 위안)보다 소폭 상향 조정되었다. 특히, 거세지는 미국과 유럽의 중국 견제 움직임 속에서 산업공급망 개선, 중국 브랜드 육성, AI 응용 확대, 과학기술 자립성 확대, 국가 전략에 부합하는 과학기술 프로젝트 진행 등을 추진하고, 이를 위해 초장기 특별국채 중 일부를 과학기술 혁신에 지원할 방침을 밝혔다. 결국, 중국 정부가 성장을 안정과 미래 성장동력 확보를 위한 '현대화 산업 건설'과 '과학기술 발전' 등의 정책 과제를 차질 없이 추진하기 위해서는, 최소 GDP 대비 -3.0% 이상의 재정적자를 상당 기간 유지할 수밖에 없다.

종합하면, 팬데믹에 대한 긴급 대응 과정에서 급격히 확대되었던 주요국의 재정적자 규모는 향후 완만한 감소 추세를 지속할 전망이다. 그러나, 인구 감소 및 고령화, 기후 변화, 공급망 재편 등 미래의 생산성과 성장률 하락을 유발하는 산적한 리스크 요인들에 대응하기 위한 기술 혁신 및 보급, 인프라 개선 등의 핵심 당면 과제를 해결하기 위해서는 통화정책보다는 재정정책의 지원이 필수적이다. 특히, 팬데믹은 주요국 모두에서 확대 재정정책의 필요성과 중요성을 일깨우는 계기가 되었다. 따라서, 국가부채 및 재정적자에 대한 우려에도 불구하고, 글로벌 금융위기 직후와 같은 급격한 재정 긴축을 통한 균형 재정으로의 복귀 가능성은 매우 낮은 것으로 예상된다. 더욱이 하반기부터 주요국 모두에서 예상되는 점진적인 금리 인하의 시작과 미국 외 국가들의 경기 모멘텀 강화는 재정 긴축의 시급성을 완화시켜주는 요인으로 작용할 전망이다. 이는 지난 2010년대 초반 미국 및 유로존에서 경험했던 급격한 재정 긴축으로 인한 경기 둔화(혹은 침체)가 재현될 가능성은 매우 제한적임을 시사한다.

**넷째, 선진국의 Tech(AI) 투자가 주도하는 생산성 제고: 1990s Redux**

지난 3월 FOMC에서 가장 흥미로운 점은 올해 미국 경제 성장률 전망치의 대폭 상향 조정(1.4%→2.1%)에도 불구하고, 핵심 인플레이션(2.4%→2.6%)과 실업률(4.1%→4.0%) 전망은 소폭 조정에만 그쳤다는 점이다. FOMC의 성장률 전망이 4Q 기준 전년 대비인 점을 감안할 때, 이는 연준이 올해 미국의 연간 성장률을 당사와 유사한 2.4%정도로 전망하고 있음을 의미한다. 또한, '25~26년에도 FOMC의 성장률 전망이 2.0%로, 잠재 수준 1.8%를 지속적으로 상회함에도 불구하고, 실업률 전망이 낮아지거나 인플레이션 정상화 예상이 달라지지 않았다. 이는 잠재 수준을 상회하는 경기 확장의 지속이 임금 및 인플레이션 상승 압력을 가중시킨다는 필립스 곡선의 결론과는 사뭇 다른 전망이다.

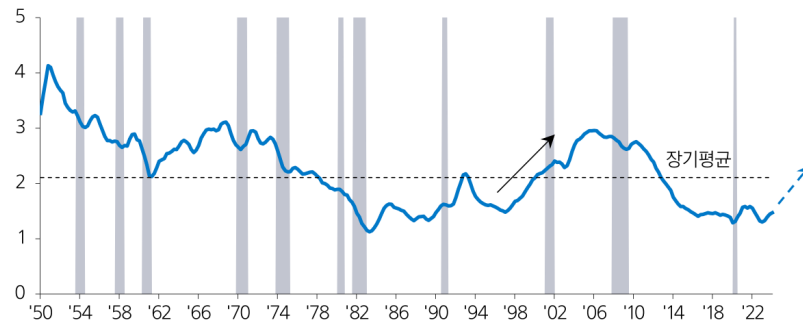
과연 FOMC의 달라진 경제 전망치는 무엇을 의미하는가? 파월 의장의 3월 기자회견에서 그 해답을 찾을 수 있다. 파월 의장은 “지난해 강력한 고용 증가세가 지속되었지만, 인플레이션 압력은 상승하지 않고 오히려 둔화되었다. 이것은 핵심 생산인구(25~54세)의 경제활동 참가율 상승과 이민 급증으로 노동 공급이 증가했기 때문이다”고 강조했다. 다시 말해, 노동 공급의 증가가 수요 측면에서 양호한 소비 및 경제 성장을 지속시킨 동시에, 미국의 총공급 능력을 확대시켜 인플레이션의 추세적인 둔화를 유발했다는 것이다. 파월 의장이 “따라서 노동시장의 공급 능력 치유(healing)에 기반한 강력한 고용 증가는 금리 인하를 지연시키는 요인이 아니다”라고 언급한 것도 바로 이런 맥락이다.

결국 지난해 미국 경제에서 관찰된 소위 “Immaculate disinflation(완벽한 디스인플레이션)”, 즉 경기 침체나 실업률 급등을 유발하지 않으면서, 인플레이션만 추세적으로 둔화된 것이 공급 능력의 개선에 기인했다는 결론에 다다른 것이다. 파월 의장은 이미 지난해 9월과 11월 기자회견에서 적어도 단기적 1~2년 정도 미국의 잠재 성장률이 팬데믹 이전의 추세 성장률보다 높아졌고, 중립금리도 단기적으로는 longer-run 2.5%보다 높아졌다는 견해를 밝힌 바 있다. 이번 경제 전망 수정에서 1) '24~26년까지 미국의 성장률이 잠재 수준을 지속적으로 상회할 것으로 예상하고, 2) 사실상 중립금리를 의미하는 longer-run dot의 median이 수년간 유지되어 온 2.50%에서 2.56%로 상향 조정된 것은, FOMC 참여자들 대부분이 팬데믹 이후 미국의 공급 능력 확대 가능성을 공식적으로 인정하기 시작했음을 보여준다.

이러한 견해는 특히, 1990년대 중반과 유사하게 AI가 주도하는 Tech 투자 붐이 생산성의 상승으로 이어질 경우, 경기 확장 사이클의 장기화와 인플레이션의 안정이 함께 진행될 수 있다는 점에서 주목할 필요가 있다. 1990년대 중반부터 미국 경제는 견고한 성장세를 보였다. 미국의 실질 GDP는 1970년대와 1980년대에 걸쳐 각각 3.23%, 3.21% 성장한 이후, '95년부터 '99년 사이 연간 4.1%의 성장률을 기록했다. 이렇게 미국 경제가 1990년대 들어 더 높은 성장률을 기록할 수 있었던 주요 요인 중 하나는 이 기간 동안 미국의 생산성이 이전보다 더 빠른 속도로 향상되었기 때문이다. '87년부터 '95년까지 미국의 노동 생산성의 개선 폭은 연평균 1.4%에 그쳤지만, 그 속도가 '95년부터 점차 가속화되면서, '00년까지 미국은 연평균 2.5%의 생산성 향상을 경험하게 되었다.

## 미국 생산성 장기 추이

(전분기 대비, %)



참고: 10년 rolling 기준. 점선은 장기평균(1950-2024년), 음영은 경기 침체 기간.

자료: U.S. Bureau of Labor Statistics, NBER, 삼성증권

어떻게 생산성이 빠르게 좋아질 수 있었는지에 대해 다양한 가설들이 있지만, 가장 주요한 요인은 미국의 민간 투자 특히, IT 부문에 대한 투자가 '95년을 기점으로 크게 증가한 점이다. 1980년대 후반부터 미국의 기업들은 대량 생산과 서비스 품질 제고를 위해 본격적으로 IT 부문 투자를 확대하기 시작했다. 1990년대 미국의 비거주자 고정투자는 7.8% 증가했는데, 이 중에서 IT 부문으로 볼 수 있는 기계장치와 소프트웨어(Equipment and software, E&S)에 대한 투자는 10.2%나 증가했다.

투자 규모가 누적됨에 따라 점차 그 속도가 빨라지면서, 기업의 IT 부문 투자는 '95년부터 '00년에 이르기까지 매년 20%씩 증가했다. 이 과정에서 컴퓨터와 같은 하드웨어뿐 아니라 소프트웨어 및 통신 장비에 대한 투자도 함께 증가하면서, 제조업을 중심으로 산업 전반의 생산성이 개선되었다. 그리고 이러한 생산성의 개선은 미국 경제의 성장으로 이어졌다. 세인트 루이스 연준 보고서에 따르면, 1980년대 미국 GDP 성장에 대한 IT 부문 투자의 기여도는 0.4%포인트에도 미치지 못했다. 그러나 1990년대에는 0.8%포인트로 그 영향이 2배 이상 증가했고, 그 중 대부분은 '95년 이후에 집중된 것으로 나타났다.

## 1950년 이후 미국 GDP 성장 기여도: '95-'00년 IT 투자 기여도 상승

GDP 성장 기여도 (%포인트)	1950년대	1960년대	1970년대	1980년대	1990년대	1995-2000년
GDP 성장률(%)	3.53	4.19	3.23	3.21	3.25	4.11
Nonresidential fixed investment	0.27	0.62	0.58	0.39	0.88	1.25
Structures	0.15	0.16	0.13	0.03	0.04	0.16
<b>Equipment and software</b>	0.12	0.46	0.45	0.36	0.84	1.09
IP equipment and software	0.06	0.17	0.27	0.32	0.58	0.78
ICT	0.04	0.14	0.21	0.30	0.55	0.75

참고: IP는 Information Processing(정보 처리 장치), ICT는 Information and communication technology(정보통신기술)를 의미

자료: St. Louis Fed

이러한 관점에서 당사는 최근 미국 설비 투자의 proxy로 사용되는 핵심자본재 신규 주문액이 지난 '22년 이후 사상최고 수준을 지속적으로 경신하고 있는 것에 주목한다. 당사는 이것이 미국이 주도하고 주요 선진국들이 뒤따르는 새로운 글로벌 Capex 사이클의 시작일 가능성이 높은 것으로 판단한다. 이는 지난 1990년대 미국 중심의 Tech 붐과 매우 유사한 성격인 반면, 2000년대 중반 중국 등 신흥국이 주도했던 Capex cycle과는 차별화된 것이다. 2000년대 중반 BRICs으로 대표되는 신흥국의 Capex는 주로 기계, 자동차/조선, 철강, 화학 등 산업재 및 소재 중심의 투자 확대였다. 이에 따라, commodity super-cycle이 필연적으로 동반되었다. 그러나 미국이 주도하는 Capex 확대는 과거 1990년대 초반부터 2000년 초까지 나타났던 것과 같이 Tech 중심의 투자 확대이다.

그러나 이번 Tech 투자 사이클을 주도하고 있는 것은 AI이다. 과거 1990년대 Tech 붐이 대부분 하드웨어 중심의 투자 확대였던 것과는 달리, 이번 AI 투자는 소프트웨어와 이를 뒷받침할 수 있는 고성능의 하드웨어 투자를 동시에 필요로 한다는 점에서 그 파급력은 훨씬 클 것으로 예상된다.

#### 미국 설비 투자의 proxy인 핵심자본재 신규 주문액 추이: 새로운 글로벌 Capex cycle의 시작

(십억 달러)



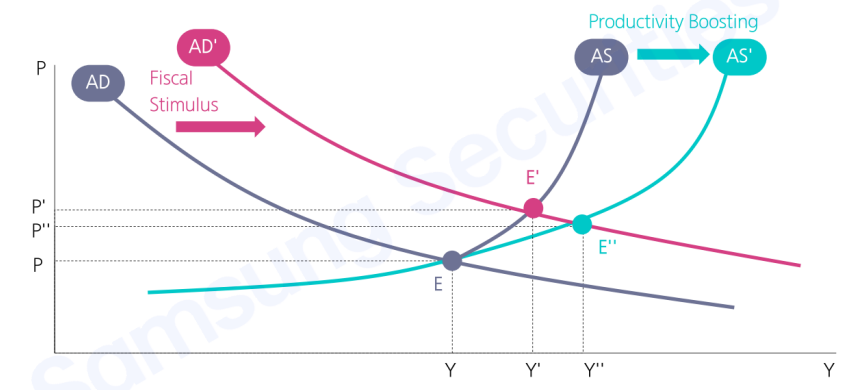
자료: Census Bureau

Tech 투자 확대를 통한 생산성의 제고, 즉, 총공급 능력의 확대가 향후 미국 및 글로벌 경제에 어떠한 영향을 미칠 수 있는지를 보다 쉽게 이해하기 위해 총수요(AD)와 총공급(AS) 곡선을 활용하여 살펴보고자 하자.

팬데믹과 이에 따른 lockdown으로 인한 갑작스러운 수요 급감에 대응하기 위해, '20~'21년 중 주요 선진국과 신흥국 모두에서 막대한 규모의 재정 확대 정책이 통화 완화와 함께 시행되었다. 이 과정에서 총수요 곡선이 AD에서 우측의 AD'로 이동하게 되었다. 일반적으로 한 국가의 총공급 능력은 단기간에 확대되기 어렵기 때문에 새로운 균형점은 기존 AS곡선을 타고 E에서 E'로 이동하게 된 것이다. 더욱이 lockdown의 반복으로 주요국의 총공급 능력이 단기적으로 축소됨에 따라, 확대 재정정책이 경제 성장률을 개선( $Y \rightarrow Y'$ )시키는 효과보다 인플레이션을 자극( $P \rightarrow P'$ )하는 효과가 더 크게 나타나게 된 것이다.

그러나 이미 누적된 AI 중심의 Tech 투자가 당사의 기대대로 향후 미국 및 글로벌 경제의 생산성 제고를 통해 총공급 능력의 확대에 이어질 경우, 이번에는 AS곡선이 우측의 AS'로 이동할 전망이다. 이는 기존의 E'가 아닌 E''가 새로운 균형점이 된다는 의미이다. 이 경우, 공급 능력의 확대로 E'에 비해 성장률은 더 높아지는 반면, 인플레이션 압력은 일정 수준에서 안정되는 골디락스 국면이 장기화될 수 있다.

#### 생산성 제고를 통한 총공급 능력의 확대는 골디락스 국면의 장기화를 의미



자료: 삼성증권

현시점에서 전 세계적인 인구 고령화를 상쇄할 수 있는 적절하고 강력한 정책 지원(이민 등)과 AI 등 새로운 기술혁신의 조합이 추세적으로 둔화되는 성장률을 생산성 향상을 통해 제고시킬 수 있는 유일한 방법이다. AI가 총요소 생산성(TFP)을 얼마나 향상시킬 수 있을지에 대한 연구는 아직 초기 단계이다. 비록 이전의 혁신적인 기술들의 보급속도와 비교하면 상대적으로 빠르지만, 절대적인 기준에서 보면 아직은 AI가 본격적으로 노동력을 대체하는 단계와는 거리가 있다. 그러나 최근 IMF가 영국을 사례를 통해 추정한 연구 결과에 따르면, 글로벌 전체적으로는 향후 10년 이상 매년 0.1~0.8%의 생산성 향상이 예상되며, 영국 등 선진국에서는 이보다 훨씬 높은 0.9~1.5%의 생산성 제고를 기대할 수 있을 것으로 추정된다. 다만, 보수적으로 가정하더라도 AI 기술적용이 본격적으로 확산되어 의미 있는 생산성의 개선이 가시화되기까지는 적어도 1~2년은 더 소요될 전망이다.

### 글로벌 생산성과 노동시장에 대한 AI의 잠재 영향

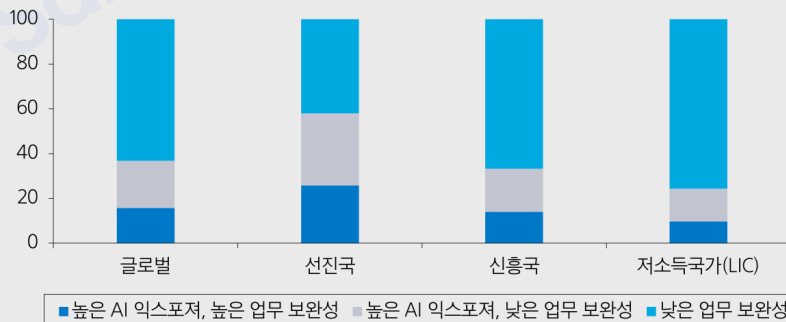
인공지능(AI)은 생산성을 향상시키지만, 일부 직업에서 인간을 대체하고 다른 직업들의 특성을 근본적으로 변화시킬 가능성이 있으며 이는 글로벌 경제에 광범위한 영향을 미칠 것으로 예상된다. 이러한 상황에서 IMF는 AI가 글로벌 경제와 노동시장에 미칠 수 있는 영향에 대해 분석했다. “AI 익스포저(AI exposure, Felten, Raj, Seamans, 2021, 2023)” 개념에 기초한 이 접근법은 AI의 업무 보완성 개념(Pizzinelli 외 2023)으로 확장되어, AI로 인해 일자리가 수해를 받거나 위험에 처할 가능성에 대한 새로운 통찰을 제공한다.

AI 익스포저는 국가별로 분명한 차이를 보이는데, 선진국은 전체 고용의 약 60%가 AI로 인한 변화에 취약하며, 신흥국에서는 40%, 그 외 저소득 국가에서는 26%가 이에 해당한다. 선진국에서 AI는 AI에 노출된 고용의 절반에 대해 생산성을 향상시키는 긍정적인 영향을 미칠 것으로 예상된다. 물론 나머지 절반의 경우, AI가 업무를 자동화함에 따라 잠재적으로 노동 수요와 임금을 감소시키고 직업 쇠퇴(job obsolescence)로 이어질 가능성이 있다. 반면, 신흥국과 개발도상국은 즉각적으로 이러한 혼란을 겪을 가능성은 낮지만 AI 도입에 따른 수해도 적을 수 있다. 많은 경우에 신흥국은 AI 기술을 효과적으로 활용하기 위한 인프라와 숙련된 인력이 부족한 상황이기 때문에, 시간이 지남에 따라 AI가 국가 간의 불평등을 악화시킬 수 있다는 우려가 제기되고 있다.

IMF는 다음과 같은 모델을 기반으로 AI가 생산성에 미칠 수 있는 잠재적 영향을 측정했다. 이 모델에서는 AI가 1) 노동 대체, 2) 업무 보완성, 3) 생산성 증가를 통해 생산성에 영향을 미친다고 보았다. 첫째, 작업의 주체가 인간에서 AI 기반 시스템으로 대체됨에 따라 작업의 효율성이 향상될 수 있다. 둘째, AI 업무 보완성이 높은 작업에 더 도움이 될 수 있다. 셋째, AI의 도입은 투자를 촉진하고 노동 수요의 전반적인 확대를 유도하면서 광범위한 생산성 증가로 이어질 수 있다. 이어지는 모델에 따른 분석은 AI 익스포저가 높으면서 가계 보유 자산 데이터를 확인할 수 있는 영국을 기준으로 하였다.

### AI 익스포저와 업무 보완성에 따른 노동시장 분포

(전체 고용 대비, %)



참고: 국가별 비중은 각국 생산가능인구를 기준으로 가중하여 계산  
자료: Cazzaniga and others 2024, ILO, IMF

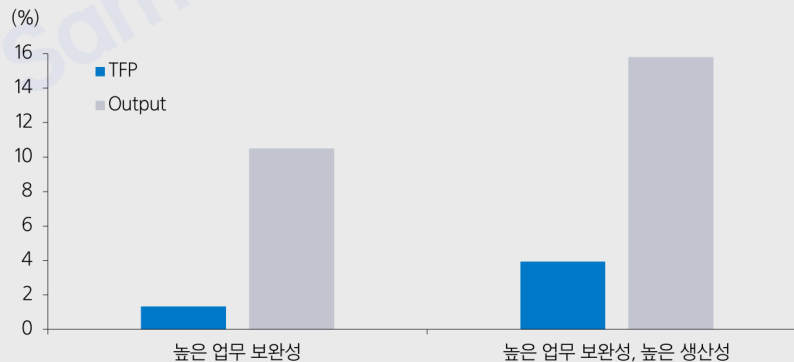
AI가 생산성에 미치는 영향은 두 가지 시나리오를 통해 분석된다. 첫 번째 시나리오(높은 업무 보완성 시나리오)에서 AI는 업무 보완성이 높은 일자리의 생산성을 크게 향상시킨다. 두 번째 시나리오(높은 업무 보완성과 높은 생산성 시나리오)에서는 AI가 전반적인 생산성을 향상시키므로써 업무 보완성을 확장하며, 이러한 직종의 생산성을 더욱 개선한다.

첫 번째 시나리오에서 AI는 자본 심화 및 소폭의 총요소 생산성(TFP) 향상을 통해 영국 경제가 안정되면서 산출량(output)이 약 10% 증가한다. 두 번째 시나리오의 경우, 생산성 영향까지 고려했을 때, output은 16% 증가하고 총요소 생산성은 4% 향상된다. 이러한 성장은 주로 과도기의 첫 10년동안 나타난다. 한편, 이 과정에서 저임금 근로자의 경우 2%, 고소득 근로자의 경우는 약 14%로 모든 근로자의 소득이 증가하면서 소득 불평등 수준은 높아지게 된다.

영국의 디지털 인프라, 숙련된 인력, 혁신 생태계 및 규제 체계를 고려할 때, AI로 인한 생산성 향상 폭은 연간 0.9~1.5% 범위일 것으로 예상된다. 반면, 많은 경우에 AI 도입을 위한 준비 측면에서 뒤처져 있는 신흥국 및 개발도상국의 잠재 이익은 영국 추정치(연간 0.9~1.5%)의 절반에 미치지 못할 것으로 예상된다. 이러한 차이의 배경은 높은 AI 익스포저 및 AI 업무 보완성을 가진 직종 종사자의 비중이 선진국 대비 신흥국에서 낮게 나타난다는 점에 있다. 선진국에서 이러한 직종의 종사자는 전체 노동자의 27%를 차지하지만, 신흥국에서는 그 비중이 16%, 저소득 국가에서는 8%로 낮게 나타난다. 이러한 차이는 AI 도입에 따른 잠재 손해가 신흥국에서 제한적일 수 있음을 시사한다.

글로벌로 볼 때, AI는 10년 이상 매년 0.1~0.8%의 생산성 향상을 촉진할 수 있을 것으로 예상된다. 그러나 AI에 따른 손해가 지역별로 큰 차이를 보인다는 점에서, AI 도입을 위한 준비가 이루어지지 않은 국가들의 여건을 개선하기 위한 국제적 협력의 필요성이 강조되고 있다. 우리는 이러한 방향에 부합하는 계획 수립을 통해 글로벌 불평등 수준을 낮추고 더 많은 국가들이 AI 도입에 따른 긍정적 영향의 대상이 되도록 할 수 있을 것이다.

#### AI가 영국 총요소 생산성(TFP)과 산출량(Output)에 미치는 영향 (시나리오별)



자료: Cazzaniga and others 2024, IMF

## '90년대 후반 미국 골디락스 경험과 주식시장 함의

### IT 버블의 짙은 그림자와 공포

IT 버블 붕괴는 주식시장 역사에 큰 트라우마를 안겼다. 2000년 3월 시장 붕괴가 시작된 이후 약 3년간 조정이 지속되었고, 이를 완전히 회복하는 데에 3년이라는 시간이 더 소요되었다 (나스닥은 무려 15년 이 걸렸다). IT 기업들의 주가 급등이 있을 때마다 버블을 우려하는 시장의 경고는 어찌 보면 당연하다.

광범위하게는 최초의 상용 인터넷 웹브라우저(넷스케이프 내비게이터)가 등장한 95년부터 '00년 초까지 이어진 IT 기업 주식의 투기열풍을 의미하지만, 일반적으로 주식시장이 'IT 버블' 하면 떠올리는 국면은 90년대 후반~'00년대 초이다. 이제는 볼 수 없는 수많은 닷컴 회사들과 우리가 익히 알고 있는 아마존('97년 5월 상장), 이베이('98년 9월) 등의 기업들이 이때 집중적으로 생겨났다.

최근 시장은 90년대 후반 IT 버블의 재출현과 '00년 초 붕괴의 재현을 우려하고 있다. 생성형 AI 때문이다. 지금까지의 가격 움직임만 보면 그 모습은 굉장히 흡사하다. 그러나 결론부터 얘기하면 현재의 AI 붐은 90년대 후반 IT 버블과 다르다. 먼저, 앞서 서술한 매크로(경기, 통화정책 등) 환경이 다르고, 더 본질적으로는 주식시장의 펀더멘털 요소에도 차이가 있다.

현재 주식시장 참여자들이 궁금해하는 건 크게 2가지이다. 첫째, AI 붐이 90년대 후반 IT 버블 경로를 따를 것인가 하는 것과 둘째, 앞서 매크로에서 주장하였듯 90년대 후반이 아닌 중반의 재현이라면 시장과 AI 주식을 어떻게 접근해야 할 것인가이다.

현 증시의 위치와 성격을 분석하기 위해 IT 버블과 AI 붐의 각 국면을 다음과 같이 정의하고, 각 국면에 시장을 주도했던 기업들을 선별하여 IT 바스켓과 AI 바스켓을 구성하였다.

### IT 버블과 AI 붐 국면 구분

- IT 버블: 1998년 1월 ~ 2000년 3월 10일 (버블 고점)
- AI 붐: 2023년 1월 ~ 현재\*

### IT 버블과 AI 붐을 대표하는 Basket 구성 종목

- IT 바스켓 (20개 기업): 델, 오라클, 쉘컴, 시스코 시스템즈, EMC, 컴버스 테크놀로지, 어플라이드 머티리얼즈, 텍사스 인스트루먼트, 자일링스, 아날로그 디바이시스, 인텔, 마이크로소프트, 테라다인, 코닝, IBM, 어도비, BMC 소프트웨어, KLA, 마이크론, HP
- AI 바스켓 (14개 기업): 마이크로소프트, 구글, 아마존, 엔비디아, 메타, 브로드컴, 오라클, 엑센추어, AMD, 램리서치, 마이크론, 시놉시스, 아리스타네트웍스, 온세미컨덕터

참고: 대표 기업 선정 기준은 각 테마와의 연관성과 당시 주가 추이 등을 종합적으로 고려  
\* 5월 23일 기준



### AI 붐은 IT 버블과 다르다

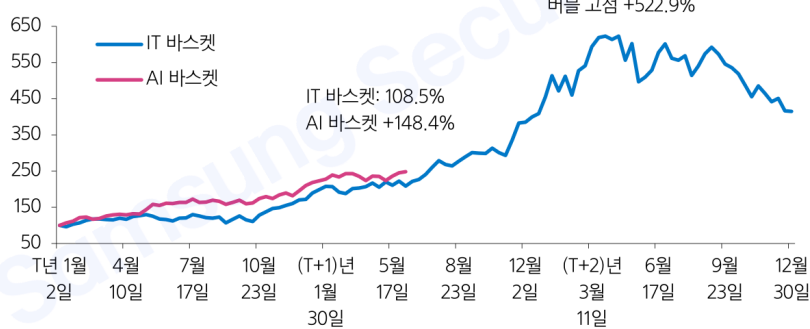
98년부터 '00년 3월 10일까지 S&P 500은 43.1%, 나스닥은 219.2% 상승했으며 IT 바스켓은 무려 동기간 518.9% 급등했다. 인터넷이라는 IT 혁신이 세상을 빠르게 바꿀 것이라는 기대와 믿음은 미국의 골디락스 환경, 통신법 개정(인프라 확대 기대) 등과 만나 버블을 계속 키웠다.

당시 장밋빛 전망이 얼마나 찬란했는지를 보려면 IT 관련 기업 IPO 규모와 CAPEX를 보면 된다. '00년 IT 업종의 연간 IPO 총규모는 700억 달러에 육박했는데, 유례없는 유동성과 이에 따른 IPO 초호황기를 맞이했던 '21년과 맞먹는 수준이다. 오히려 20년 전임을 감안하면 훨씬 능가한다고 할 수 있다. CAPEX도 마찬가지이다. 90년대 후반 IT와 커뮤니케이션 업종의 CAPEX 규모는 현재 수준과 비슷하고, S&P 500 전체 대비 비중으로 따지면 무려 70% 이상이 IT 관련 투자였다.

산업/생활 전반에 가져올 혁신적인 변화에 대한 기대, 이를 보여주는 폭발적인 주가 상승과 주요국/빅테크의 막대한 반도체 투자까지, 이러한 요소들은 AI 붐과 IT 버블이 꽤 닮아 있음을 보여준다. 특히 대부분의 신기술이 캐즘을 반드시 넘어야 하고, 그 과정에서 노이즈를 피하기 어렵다는 점에서도 그러하다. 90년대 중반 상용 인터넷이 출시됐지만 인프라와 하드웨어, 소프트웨어가 실제로 뒷받침되기까지는 시간이 걸렸고, '00년대 중반이 되어서야 인터넷은 보편화될 수 있었다.

### IT 버블을 연상케 하는 AI 주식들의 상승 속도

(지수화: T년 1월 = 100)



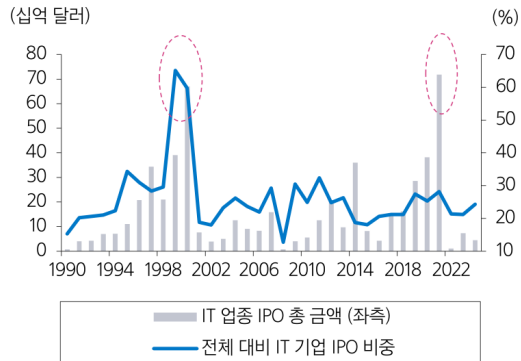
참고: T년은 각각 1998년과 2023년, 바스켓 가격은 단순 평균, 5월 23일 기준  
자료: LSEG, Factset, 삼성증권

버블 붕괴는 기대와 현실이 어긋나는 데에서부터 시작한다. 그 현실을 가장 객관적으로 확인할 수 있는 지표는 기업 실적이다. 당시 수많은 인터넷 기업들은 소비자에게 효용을 주고 실적을 증명하는 데에 실패했다. 기대감으로 상승한 주가는 멀티플만을 끌어올렸고, 그 거품은 '99년 말 금리 인상과 함께 무너졌다. 반면 AI는 다르다. 실적을 계속 보여주고 있다는 점이 가장 중요하고도 근본적인 차이이다. AI 가격이 비싸다고 지적하면서도 IT 버블과 다를 수 있다고 인정하는 시선이 '23년보다 늘어난 것도 이 때문이다.

물론 초기 투자 수요 단계에서 나오는 폭발적인 이익 성장률은 필연적으로 둔화될 수밖에 없다. 시장의 시선이 실제 효용에 대한 의구심과 reality check로 이동하고 있다는 점 또한 AI 주가 모멘텀이 이전과 같은 속도로 움직이기 어려움을 시사한다. 이미 시장은 그 구간에 진입하기 시작했을 수도 있다. AI 관련 호재에 일제히 주가 상승을 같이 하던 AI 바스켓 종목들이 이제는 조금씩 방향이 차별화되고 있고, 단순히 실적이 좋으면 가격을 올렸던 시장은 그래서 그 다음은 무엇인가에 대해 질문을 던지고 있다.

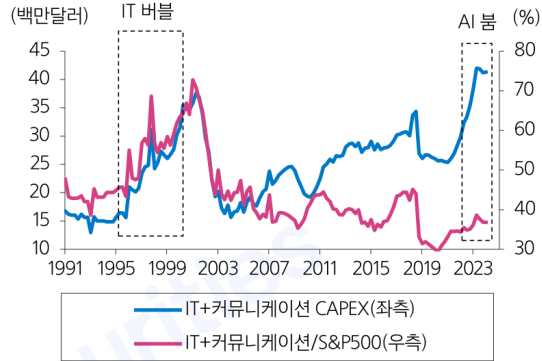
하지만 지금까지의 주가 상승이 단순 기대에만 의존하지 않았다는 점에 주목해야 한다. 이는 적어도 향후 주식시장의 위험 요인이 발생한다 하더라도 그 충격이 IT 버블과 다를 가능성이 매우 높음을 보여주기 때문이다. 올해도 금리 노이즈가 계속되며 채권금리가 재차 상승하고, 변동성이 높아진 상황에서도 AI 바스켓 종목들이 오른 가격 대비 경미한 조정, 상대적으로 견조한 주가를 보인 것은 이를 뒷받침한다. '00년 초 IT 버블의 촉매가 유동성 축소(금리 인상)이었음을 생각하면 올해 하반기 이후 환경에서 이 위험을 걱정할 필요는 적다.

## IT 기업 IPO 추이



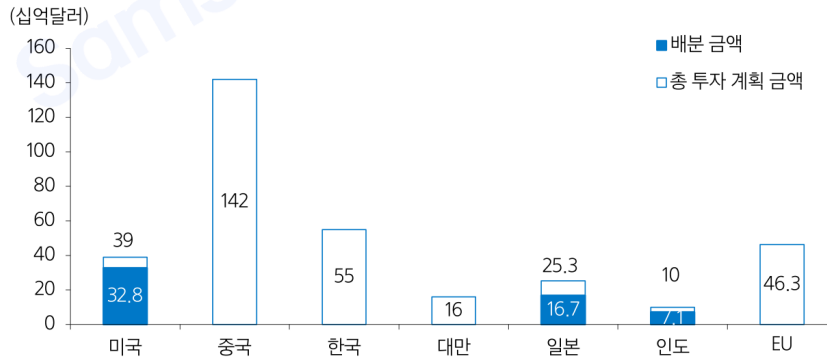
참고: 인터넷, 정보통신, 소프트웨어, 반도체, 미디어, 광고, 컴퓨터 업종 대상  
미국 거래소 상장 기준, SPAC은 제외  
자료: Bloomberg, 삼성증권

## IT+커뮤니케이션 CAPEX 추이



참고: S&P 500 기준  
자료: Bloomberg, 삼성증권

## 글로벌 반도체 칩 투자 계획 및 현황



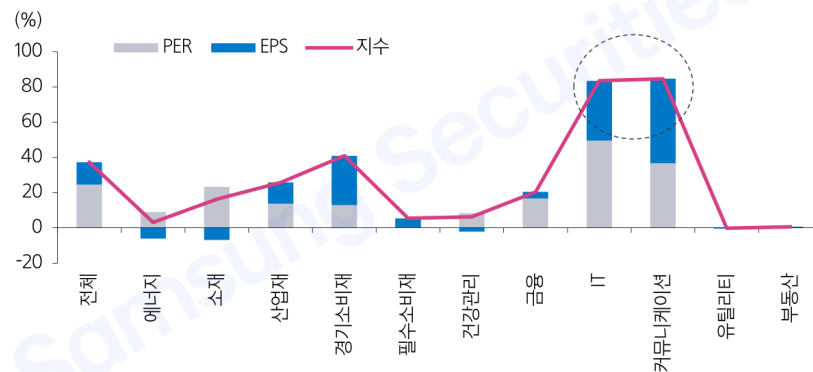
참고: 각국 정부에서 발표한 보조금, 세제 인센티브 기준, 중국의 경우 투자펀드 조성 계획 규모  
자료: Bloomberg reporting and research, Semiconductor Industry Association, 삼성증권

## 실적이 뒷받침해주지 못했던 IT 버블 국면



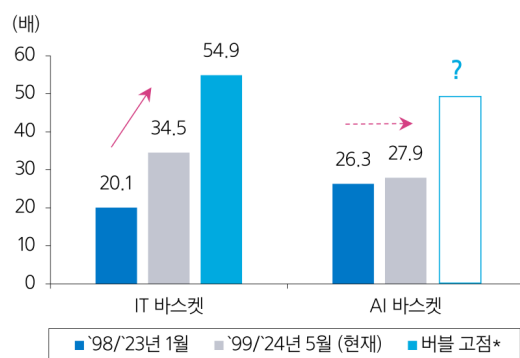
참고: 12m fwd EPS 기준, 5월 23일 기준 / 자료: LSEG, 삼성증권

## '23년 이후 S&amp;P 500 업종별 성과 기여도 분해

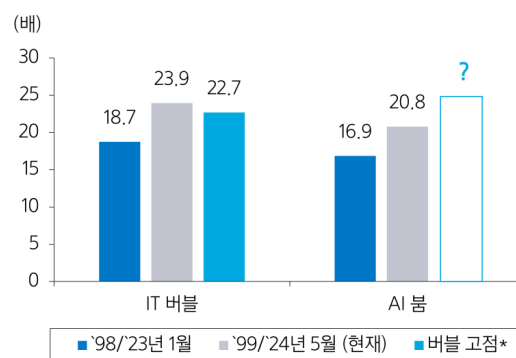


참고: '23.1월부터 5월 현재까지 수익률 기여도 분해, 12m fwd 기준, 5월 23일 기준 / 자료: Factset, 삼성증권

## P/E 밸류에이션 - IT vs AI 바스켓

참고: 5월 23일 기준, \* '00년 3월 10일 기준 12m fwd P/E  
자료: LSEG, Factset, 삼성증권

## P/E 밸류에이션 - S&amp;P 500 (동 기간)

참고: 5월 23일 기준, \* '00년 3월 10일 기준 12m fwd P/E  
자료: LSEG, Factset, 삼성증권

**불균형은 언젠가 균형을 찾는다**

두 번째, 시장 참여자들의 다음 고민은 버블이 아니라면 이러한 AI 주도 시장 랠리가 더욱 지속될 수 있을까? 이 구도가 전환되기 위한 조건은 무엇일까? 이다. 이는 단기적으로는 올해, 중기적으로는 2~3년 시장 투자 전략에 대한 일부 답이 될 것이다.

풍부한 현금 흐름을 바탕으로 주요 빅테크 기업들은 예상보다 더 빠르고, 공격적으로 AI 기술 발전을 이끌고 있다. 시장이 이를 인정하면서도 한 권으로 부담스러워하는 이유는 소위 M6 (테슬라 제외), 매우 소수의 기업을 중심으로 주식시장이 계속 움직여왔기 때문이다. 몇몇 저렴한 업종/주식/스타일로 이동하려는 시도가 있었지만 매크로/지정학 불확실성, AI 관련 호재 등에 매우 민감해, 그때마다 우위 성과를 쉽게 반납하곤 했다.

사실 AI 랠리에 가려진 현 주식시장의 문제는 단순히 AI 주식 가격이 비싸다는 것이 아니다. 본질적으로는 시장 집중 현상(Market concentration)이 너무 과도하다는 것이 문제이다. 1) 시총 상위 10개 기업으로의 쏠림, 2) IT 업종으로의 쏠림, 3) 대형주로의 쏠림이 지속되고 있고, 이러한 현상은 AI로 인해 최근 더욱 강화된 모습이다.

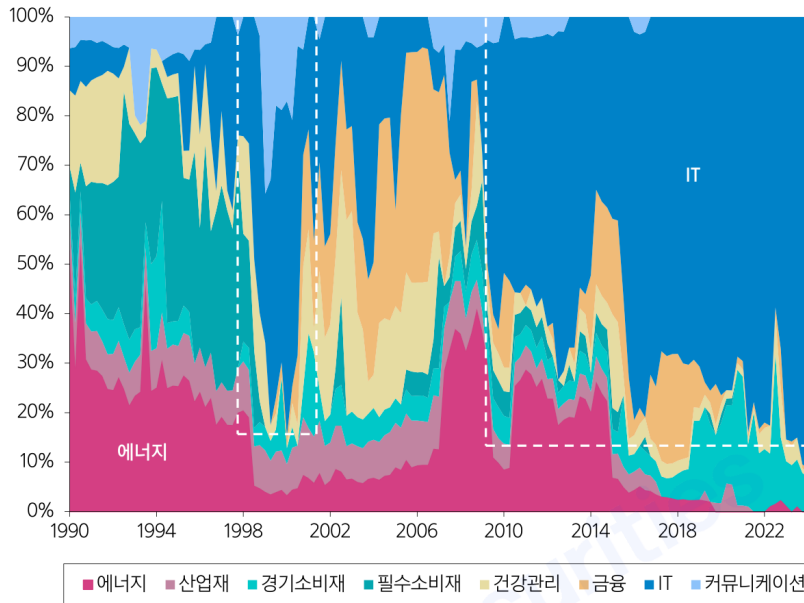
상위 10개 기업과 11~30개 기업 시가총액 간 갭은 30년래 최고 수준이다. 상위 10개 기업의 업종 지배력을 보면 IT 집중 현상은 더욱 극명하게 드러난다. 상위 10개 기업에서 차지하는 IT 업종의 영향력은 90%에 육박하고, 그 다음으로 비중이 높은 경기소비재조차 결국 아마존임을 고려하면 사실상 그 성격은 100%에 가깝다.

역사적으로 이러한 쏠림 현상은 어떤 방식으로든 균형을 찾았다. 대개 많이 오른 주식이 떨어지거나 오르지 못한 주식이 오르거나이다. 전자라면 향후 주식시장은 안전자산 선호로 돌아감을 의미하고, 후자라면 위험 선호가 계속되는 환경을 의미한다. 미국의 골디락스, AI 주식은 물론 이외 기업들의 실적 펀더멘털 등을 고려할 때 이번 국면은 후자가 될 것으로 전망한다.

**시장 집중(market concentration)은 30년래 최고 수준**

참고: S&P 500 전체 시가총액 대비, 5월 23일 기준  
자료: LSEG, 삼성증권

## S&amp;P 500 시총 상위 10개 기업의 업종 지배력



참고: 업종분류는 ICB(Industry Classification Benchmark) level 1 기준, 커뮤니케이션(Telecommunication)은 통신서비스만 포함  
상위 10개 종목의 업종별 시장 영향력을 보기 위해 시가총액 가중치를 반영  
자료: LSEG, 삼성증권

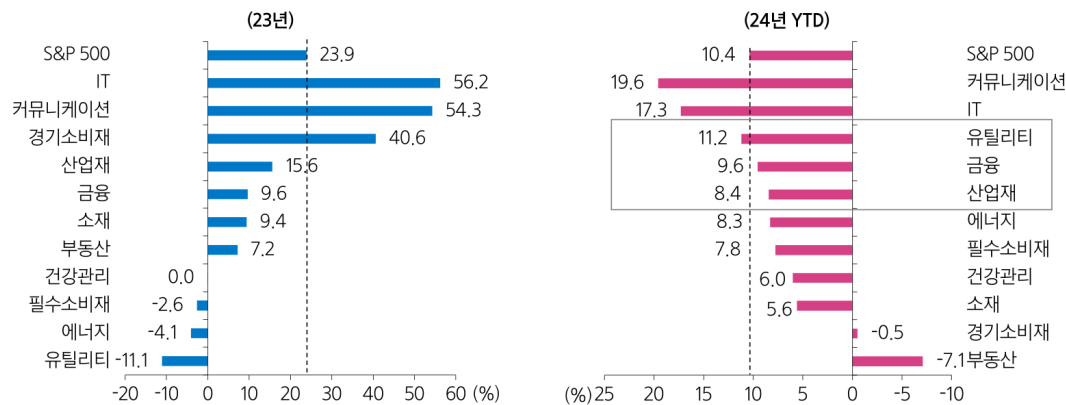
올해 주식시장의 시선이 조금씩 점차 다른 업종으로 이동/확대되고 있는 모습이 관찰되고 있다. 지난해 S&P 500 연간 성과는 23.9%로, 시장 지수를 아웃퍼폼한 업종은 IT, 커뮤니케이션, 경기소비재뿐이었다. 올해도 IT, 커뮤니케이션의 상승이 가장 두드러지지만 이외 유틸리티, 금융, 에너지 등이 시장 지수를 아웃퍼폼하고 있다.

그동안 저렴한 Non-AI 주식으로의 시장 확산이 쉽지 않았던 이유는 매크로 불확실성이 계속되어왔던 것이 주효했다. 금리 인상이 지속되고, 고금리 환경이 지속되는 등 유동성이 제약적인 환경에서는 결국 실적이 좋고, 현금 흐름이 풍부하며 AI 테마를 이끌고 있는 빅테크가 유일한 대안이자 최선의 선택이기 때문이다.

그렇다면 달리 말해, 하반기 시장이 금리 인하에 대한 기대를 회복하고, 견조한 경기가 금리 인상을 자극하지 않을 것이란 기대와 자신감을 되찾아갈수록 시장의 시선은 다른 업종/스타일 주식으로 확대될 개연성이 높다. AI를 중심으로 한 IT 기업들의 주도력이 완전히 전복됨을 주장하는 것이 아니다. 심화되었던 시장 쏠림이 하반기로 갈수록 일부 완화될 수 있는 환경이 다가오고 있음을 의미한다.

이를 지지할 또 다른 조건은 실적 모멘텀의 변화이다. 지난해부터 지속된 AI 밸류체인이 이익 성장은 점차 둔화될 것으로 전망되는 반면, 다른 업종의 실적 모멘텀은 개선될 것으로 예상되고 있다. 대표적으로 S&P 500 반도체 실적은 전년 대비 24년(E) +35.8% → 25년(E) +29.7% → 26년(E) +13.5%로 둔화되고, 구글/메타 등이 속한 종합 미디어&서비스는 전년 대비 24년(E) +30.9% → 25년(E) +13.4% → 26년(E) +14.2%로 둔화될 전망이다. 따라서 그간 가격이 계속 부진했고, 향후 실적 개선세가 긍정적인 업종/주식으로 시선을 확대해 나아감이 적절해 보인다.

## S&amp;P 500 업종별 성과 리뷰



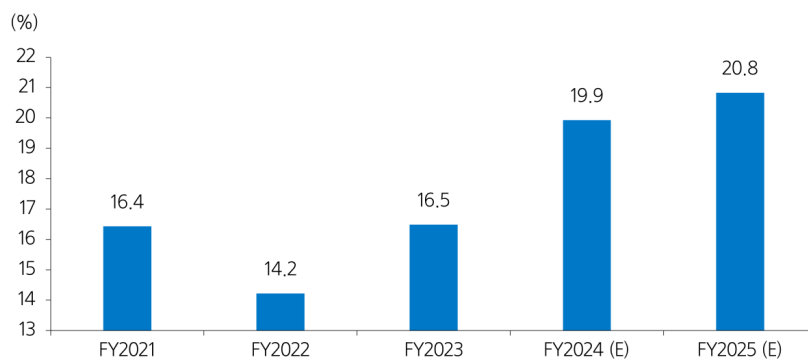
참고: 5월 23일 기준 / 자료: Factset, 삼성증권

## S&amp;P 500 업종별 실적 증가율 전망

(%, 전년 대비)	2023	2024	2025	2026
S&P 500	1.99	11.37	14.04	11.58
에너지	-26.47	-3.99	10.03	4.73
소재	-23.27	-0.88	15.05	10.28
산업재	20.31	7.22	15.88	15.71
경기소비재	47.89	14.32	15.66	16.52
필수소비재	2.73	5.09	7.82	8.09
건강관리	-20.70	8.96	18.32	10.74
금융	2.47	12.78	11.09	12.51
IT	6.40	17.78	17.79	7.72
커뮤니케이션	29.31	22.20	12.85	13.02
유틸리티	5.00	8.27	8.37	7.86
부동산	0.97	0.95	6.35	7.58

참고: 12m fwd EPS 기준, 5월 23일 기준 / 자료: Factset, 삼성증권

## M6의 S&amp;P 500 전체 이익 대비 비중

참고 S&P 500개 순이익 합산 기준, 1~3월 결산기업의 경우 FY24는 확정치, 테슬라 제외, 5월 23일 기준  
자료: Factset, 삼성증권

## Tech: 반도체 지금 더 사세요

“성공 비결은 3번 생각하는 습관. 대학 시절 ROTC였는데 포탄을 쏠 때, 길게 한 번 쏘고, 짧게 한 번 쏘고, 그다음에 제대로 쏜다는 것을 배웠다. 무언가를 결정할 때 길게 한 번 보고, 짧게 한 번 보고, 그다음에 제대로 결정하는 것이다.” (워런 버핏과 함께 버크셔 해서웨이를 이끈 찰리 멩거)

당사의 매크로 전망은 미국 및 글로벌 경기 확장 사이클이 장기화될 것으로 예상하였다. 하지만, 멩거의 의견과 같이 단기적인 관점도 확인이 필요하다. 높은 Valuation과 이익 전망치는 부담되고, 경기 불안으로 기업의 투자가 위축될 수 있다는 우려가 있기 때문이다.

**이익 전망치 상황. 누가 기술과 시장을 주도할까?** 당사의 반도체 사이클 (2025년까지) 전망은 긍정적이다. 삼성전자와 SK하이닉스의 이익 전망치를 상향하고, 이는 주가가 더 오를 수 있다는 의미이다. SK하이닉스가 여전히 Top Pick이지만, 기대감이 최저인 삼성전자에 대해서 12개월 목표가를 11만원에서 12만원으로 상향 조정한다. ‘좋은 회사’의 조건(높은 마진과 현금흐름, 유지 가능한 경쟁력)을 구비한 회사도 더욱 투자하자. 메모리 공급망에서 보면, 인프라의 엔비디아, 클라우드와 서비스의 마이크로소프트와 테슬라가 기술과 시장을 주도하고 있다.

구체적으로 이 리포트는 다음과 같은 세 가지 질문에 대한 답변이다.

**인공지능 시장의 규모는 시장의 기대치를 넘어설 것이다.** 첫째, 인공지능 수요에 대한 질문은 기술을 아는 것이 우선이다. 기술 방향이 반도체 사용량을 키워가고 있다. AI 거대화 모델은 더욱 빠르고 많은 메모리가 필요하다. On-device에서도 하드웨어 대역폭의 한계는 교체 수요를 자극할 것이다. 둘째, 공급 환경은 척박한 상황이다. AI 반도체를 제조하기 위해 낮은 수율과 커지는 사이즈가 문제이고, 기업들은 투자에서 Agility를 강조하고 있다. 하반기와 내년 반도체 시장은 공급부족으로 시장의 이익전망 상향 가능성이 높다는 판단이다.

**인공지능의 혁신은 상상 이상일 것이다.** 디지털 전환은 VHS 테이프를 CD로 전환하기 보다 넷플릭스의 탄생에 의미한다. IT 혁신은 생산성의 가치이고, 제조 효율을 넘어 신약 개발은 인구 피라미드의 변화와 반도체 시장의 성장을 의미한다. 핵심은 트랜지스터의 혁신이고, 연산 컴퓨팅은 물리적인 공간과 융합하며 로봇과 자율차 시장으로 이어진다. 인공지능의 승자는 과거 노동력과 제품을 수출하던 것을 넘어 사회 시스템을 수출하는 강력한 성장을 차지할 것이고, 성장의 열매는 하드웨어와 소프트웨어를 동시에 최적화하는 기업에게 돌아갈 것이다.

**긍정적인 전망에 버금가는 Risk도 상존한다.** 단기적으로 2000년 IT 버블 붕괴와 같은 상황을 예상하지 않으나, 문제는 AI가 마케팅 구호로 그쳐 소비자 만족을 얻지 못하는 것이다. 이메일을 제공하는 인터넷 초기 제품과 달리 소비자는 멀티미디어를 원했던 것과 같이, AI 서비스도 비용대비 효율이라는 관점에서 소비자의 외면을 받을 수 있다. 동시에 AI를 기점으로 중국발 첨단제품의 공급 과잉이 시작될 수 있다는 생각이다. 단기적으로는 생산된 제품을 국내 소비하기 위한 정책이 수요를 이끌 수 있으나, 장기적으로 기존 공급사에게 위협이 될 수 있다.



**AI, Big thing is coming**

**벌거벗은 임금님.** 인공지능(AI, Artificial Intelligence)에 대한 기대와 우려는 극과 극을 달리고 있다. 당장 세상을 바꿀 것 같기도 하고, 기술의 진보는 인정하지만 가시적인 변화는 느리고 오히려 단기적으로는 인공지능의 열기가 식으며 또 하나의 IT 버블이 터질 것이라는 불안감도 존재한다. 때문에 90년대 초의 PC와 컴퓨팅 혁신에 비교되기도 하고, 90년대 말의 인터넷 버블에 비교되기도 한다.

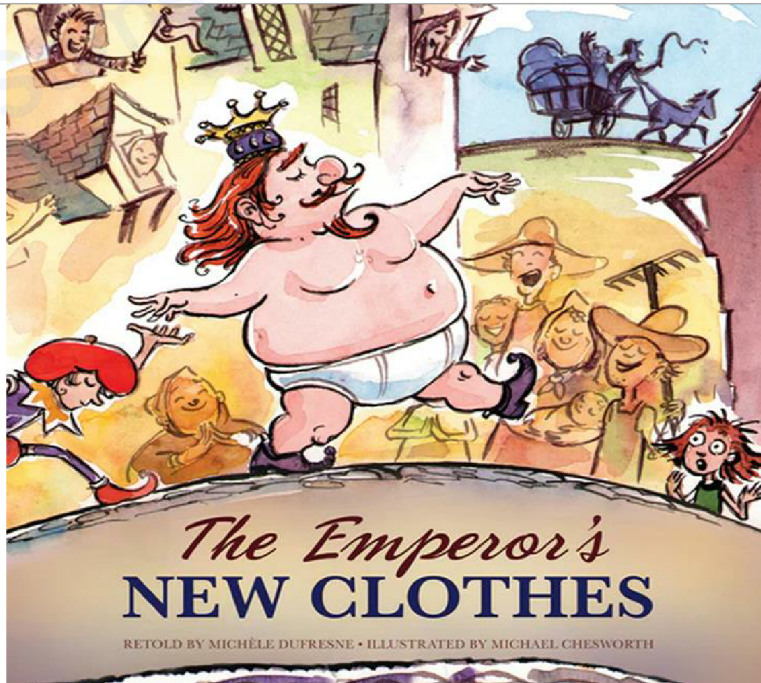
안데르센의 동화 '벌거벗은 임금님'에서 모두가 보는 벌거벗은 임금님을 정작 본인은 보지 못하고 있는 것처럼, 많은 기업가와 투자자들이 새로운 'Gen-AI'라는 실제로는 없는 옷을 입고 좋아한다는 시니컬한 비평도 있다(The High Tech Strategist, 2024년 5월 1일).

**중요한 질문들.** 기술에 관련된 질문은 기술을 이해하는 것이 혼란을 줄이는 지름길이다. 하지만 인공지능에 투자하는 기업과 투자자에게 더 필요한 것은 쉬운 질문에 답하는 것이다. 인공지능 기술이 어떤 가치를 제공하며, 어떻게 우리 현실 사회에 영향을 주는지, 기업들은 어떻게 여러 가지 어려움을 극복하고 수익모델을 창출하는지, 또는 이러한 기대감이 물거품이 될 수 있는 위험 요인은 무엇인지 살펴보는 것이다.

기술과 함께 시장에 대한 질문에 대해서도 답해야 한다. 과거 PC나 인터넷 컴퓨팅의 혁신이 미국을 중심으로 전개되었다면, 인공지능의 혁신은 미국과 중국, 유럽과 아시아 모두가 도전하는 글로벌 현상이다. 때문에 미국과 중국 사이에 벌어지고 있는 갈등과 향후 영향을 예측하는 것도 중요하다.

**기술과 시장을 주도하는 회사들.** 엔비디아, 마이크로소프트, 테슬라, 그리고 메모리. 결론은 반도체 주식을 더 사야 한다는 것이다. '좋은 회사'(높은 마진과 현금흐름, 유지 가능한 경쟁력)를 더욱 투자해야 한다. 메모리 공급망에서 보면, 인프라의 엔비디아, 클라우드와 서비스의 마이크로소프트와 테슬라가 기술과 시장을 주도하고 있다. 또한 메모리 주식도 하반기 추가 상승이 예상된다. SK하이닉스와 함께 특히 시장을 오랫동안 Underperform한 삼성전자를 이제는 적극 매수해야 할 때라는 판단이다.

**벌거벗은 임금님은 AI라는 새로운 옷을 입었을까?**



자료: 구글이미지, 삼성증권



**인공지능 수요 성장, 기술을 알면 덜 불안하다**

인공지능 기술의 혁신은 우리가 경험했던 어느 사이클보다 길게 유지될 것으로 판단한다. 브릭스(BRICS)의 부상과 Commodity 사이클이 길어도 5년 정도였다면, 과거 PC와 스마트폰의 혁신은 짧게는 10년, 또는 그 이상 지속되었다. 지금을 과거와 비교할 수 있다면 인터넷 버블을 경험하던 90년대 말보다는 PC 기술이 확대되던 90년대 초반이나 중반이 어울린다.

물론, 사이클은 피할 수 없다. 반도체 주가가 작년 초반 이후 랠리를 보인 이후 금리에 대한 불확실성이 커지며 시장의 변동성이 커지고, 따라서 AI 사이클의 단기 고점 논란이 발생하는 것은 불가피하다. 금리가 오르는 환경이라면 기업들이 투자를 줄일 수 있기 때문이다.

하지만, 당사는 올해, 그리고 적어도 내년까지 AI를 배경으로 IT 산업이 호황을 이어갈 것으로 판단한다. 그 근거를 설명하기 앞서, 우선 독자의 이해를 돕기 위해 기술적인 배경을 설명할 것이다. (p. 64 부록 1. 인공지능 기술 개요 참고)

**클라우드 AI 기술은 최적화 대신 거대화를 선택. 더 빠르고 더 많은 메모리는 필수**

**장님 코끼리 다리 만지기 같은 수요 전망.** 인공지능의 반도체 수급에 대한 불안감은 수요 전망이 불확실하기 때문이고, 이는 수요가 알기 어려운 소위 블랙박스 형태로 진행되기 때문이다. 2016~2018년 클라우드 확대에서 경험한 바와 같이 어느 고객이 어느 정도 주문을 할지는 스마트폰 판매 수량과 탑재량을 계산하는 것과 달리 어렵다. “장님 코끼리 다리 만지듯이” 수요 전망을 하는 것만으로는 수요를 장담하기 어렵기 때문에, 거시경제의 문제가 발생하면 당장 공급 과잉이 발생할 수 있다.

**The bigger is the better.** 그러나 올해 현재까지 인공지능의 수요(TAM, Total Available Market)는 계속 상향 조정되고 있다. 그 배경은 클라우드에서 더욱 많은 연산으로 더욱 똑똑한 결과를 얻어내기 위해 동시에 구동되는 인공지능에 있다.

이 같은 과정은 학교 때 우리가 교과서보다 문제집을 많이 풀며 점수를 높여가는 과정과 유사하다. 정확도를 높이기 위해 더 많은 데이터를 돌리는 것이 초거대 모델이기에 더 많은 반도체 탑재량이 필요하다. 거대화의 방향성은 2018년 말 클라우드 확대에서 경험하였던 서버 시스템의 효율화 (Software Optimization)를 통해 투입되는 반도체 탑재량을 줄여가는 것보다 2016년 클라우드 확대의 초기와 유사하다.

**거대화 모델을 위해 더 빠르고 더 많은 메모리는 AI의 필수 조건.** 파라미터를 늘리고 모델을 거대화하여 똑똑한 대답을 하는 AI를 얻어도, 대답하는 속도가 느리면 안된다. 우리가 GPT에게 “최근 읽어볼 만한 AI 리포트가 뭐야?” 질문을 하고, GPT가 “삼성증권의 5월 AI 리포트를 추천합니다.”라고 답한다고 가정해보자.

GPT는 학습한 데이터를 모두 활용한 뒤 ‘삼성’이라는 단어를 쓰게 된다. 하지만 이 같은 상황은 ‘증권’이라는 단어를 쓸 때도 변하지 않는다. 때문에 데이터를 전송하는 대역폭이 충분히 넓지 않다면, 데이터가 이동하는 시간 동안 프로세서는 놀게 되고 서비스는 느려지게 된다.

엔비디아와 같은 GPU 업체에게 메모리의 용량을 키우고 속도를 높이는 것이 만족할 만한 AI를 구현하는 첫걸음이다. GPU가 값싼 기존 범용 디램 대신에 수율도 낮고 가격도 비싼 HBM(High Bandwidth Memory, 고대역폭 메모리)을 고집하는 이유다. 아직 수요의 정체를 고민하기에는 이르다는 뜻이다.

**초거대화가 유발하는 열과 전력 소비. 그리고 따라오는 수요 성장과 공급 제한**

**열나고 데이터센터 공간과 전기 잡아먹는 하마, GPU.** 연산량을 더 키우다 보면 일반 서버를 CPU가 탑재된 AI 서버로 교체하는 것이 필요하다. GPU는 열이 많이 나서 조밀하게 설치하지 못하고, 동시에 전력 소비가 크다. GPU를 위한 공간이 많이 필요하고, 스토리지 등에서 공간과 전력 소비를 줄여야 한다. 더 작아진 스토리지를 위한 공간에는 기존에 사용하던 HDD를 낸드로 교체하게 된다. 고용량 낸드를 사용하면 공간 문제와 함께 최근 AI에서 가장 중요한 문제로 부각된 전력 소비를 줄여주기도 한다.

**갑자기 부상한 기업용 QLC 낸드.** 최근 우리는 예상하지 못했던 두 가지 수요의 증가를 목격했다. 하나는 강력한 추론(Inference)에 필요한 고용량 QLC 낸드(4 bit per cell)의 등장이고, 두 번째는 보다 소비 전력을 줄인 HBM이다.

AI를 구동하는 GPU는 바로 옆에 이를 지원하는 HBM이 탑재된다. 하지만 빠른 속도와 대역폭을 제공하는 디램과 달리, 데이터를 저장하는 스토리지는 AI와 거리가 멀고 용량을 저가에 높이는 것일 뿐이라는 시각이 지배적이었다. 하지만 올해초부터 일부 고객은 강력한 추론을 위하여 보다 많은 GPU를 탑재하고, 이 경우 제한된 전력이 문제가 된다. 따라서 기존 저장 장치인 HDD만큼 많은 고용량 저장 공간을 확보하는 동시에 보다 소비 전력을 최대한 제한할 수 있는 QLC 낸드가 부각되고 있다.

문제는 QLC 낸드의 신뢰성이다. 기존 낸드 시장의 80% 이상을 점유하는 TLC (3 bit per cell)에 비해 QLC는 더 짧게 신호를 켜고 읽어야 하므로 신뢰성이 떨어지는 물리적 현상을 벗어나기 어렵다. 대부분의 기업 수요가 아직 CTF 기술 기반의 QLC를 수용하지 않는 이유다. SK하이닉스가 인텔의 낸드를 인수하여 확보한 Floating gate 기술기반 QLC 수요가 추론용으로 늘고 있고, 삼성전자는 CTF 기술이지만 Software 기술을 활용하여 QLC의 신뢰성을 TLC급으로 개선하는 것으로 알려져 있다. 하반기에는 이러한 추세가 일부 On-premise 고객을 넘어 하이퍼스케일 고객으로 확산될 것으로 예상되고, 이는 낸드 수급과 가격 변화에 긍정적이다.

**더욱 까다로워진 HBM 설계와 생산 환경.** 또한 차세대 HBM의 품질 조건에서 전력 소비가 더욱 중요하게 부상하고 있다. 전력 소비를 줄이려면 메탈 배선의 수를 늘리거나, 고속의 데이터 이동을 지원하는 물질 개선 등 더 까다로운 설계와 생산환경이 필요하다. 이는 AI 반도체의 공급 과잉이라는 우려를 상당히 희석하는 요소라는 판단이다.

**솔리다임 NAND의 HDD 대체 효과**

자료: Solidigm

**온디바이스 AI의 목표는 소프트웨어 최적화, 그리고 충분한 대역폭을 위한 하드웨어 교체 수요**

온디바이스 AI는 비용 절감이 우선 목표. 온디바이스 AI를 추구하는 것은 개인의 민감한 데이터를 클라우드에 저장하지 않는 것 보다는 비용 절감 측면이 강하다. 수익명이 한꺼번에 클라우드에서 AI를 실행할 때 발생하는 클라우드 비용이 너무 크기 때문이다. 서비스 제공자 입장에서 이러한 비용을 소비자에게 전가하기 어렵기 때문에 각자의 디바이스에서 어느 정도의 AI를 실행하여 비용을 줄이고자 하는 것이다.

**초거대 모델(LLM)을 스마트폰에서 구동하면 어떤 일이 발생할까?** 우선 실행 중인 애플리케이션들이 느려지거나, 또는 LLM이 만족할 만한 성능을 내지 못할 수 있다. 만족할 만한 성능을 보장해도 과열되거나 배터리 소모가 클 수 있다. 때문에 LLM을 경량화하는 작업을 하고 있다.

이 작업들은 시간이 필요하지만, 스마트폰 제조사들은 기다리기보다 경쟁적으로 LLM을 탑재하고 있다. 이는 온디바이스 AI가 최적화라는 방향으로 진행되고 있지만, 역설적으로 스마트폰 화질 개선과 카메라 등 성능 개선으로 매년 새로운 스마트폰을 기다렸던 때로 시간을 되돌릴 수도 있다고 판단된다.

**온디바이스 AI가 불러올 스마트폰 교체 수요.** 애플이 하반기에 LLM에 필요한 연산을 줄일 수 있다면, 그것도 열과 소비 전력 문제를 동시에 해결한다면 온디바이스 AI는 완전히 다른 양상으로 진행될 것이다. 또한 이는 대부분의 스마트폰의 교체 주기가 다시 빨라진다는 것을 의미한다.

**스마트폰이 제공하는 메모리 대역폭과 LLM이 요구하는 대역폭을 비교하여 보자.** 아이폰 기준 메모리 대역폭(Bandwidth)은 초당 30~50GB 수준이다. 아이폰 15는 LPDDR5 6GB의 용량과 50GB/sec 수준의 대역폭을 제공한다. 6.4GHz로 구동되는 LPDDR5 x 64비트 / 8 (1 바이트 당 8 비트)를 계산하면 51GB가 나온다. 중저가 AP가 32비트로 구동되고, 4.2GHz로 구동되는 LPDDR4이라면 대역폭은 더욱 떨어진다.

여기에 양자화 기술을 도입하여 파라미터당 4 bit 수준으로 경량화한 LLM을 적용한다고 가정하자. 그러면 마이크로소프트 Phi-3-mini는 초당 하나의 토큰을 생성하는 데 필요한 대역폭이 1.8GB, 메타 Llama의 경우 3.5GB가 필요하다.

실제 LLM을 구동할 경우 아이폰 15에서는 초당 19개의 토큰을 생성할 수 있다고 하고, Phi-3-mini의 경우 LLM 이외 다른 앱을 구동하는 것을 감안하여 12개의 토큰을 생성한다고 한다. 토큰은 AI 학습을 위한 말뭉치라고 할 수 있다. 예를 들면 삼성증권이라면 삼성, 증권이라는 2개의 토큰을 사용하여 의미를 부여하는 것이다.

12개의 토큰이 가능하다면 초당 4개 수준의 단어를 생성하는 것이다. 초당 4개의 단어를 대답한다면 다소 느리다. 때문에 업계는 LLM이 하나의 Token을 생성하기 위한 대역폭을 초당 1.8GB에서 1GB 수준으로 줄이려고 노력하고 있다. 줄일 수 있다면 초당 8개의 단어를 대답할 수 있다. 트랜스포머 모델을 기반으로 초거대 모델이 시작되었지만, 계속 추론을 늘리다 보면 너무 많은 용량이 필요한 트랜스포머 모델의 문제점을 개선하는 것이기도 하다.

**문제는 공급이야****2025년 생산 물량도 이미 대부분 판매가 완료**

**추가 증설은 고객의 추가 주문이 전제 조건.** SK하이닉스는 5월 언론을 통하여 “내년 HBM 생산량은 이미 대부분 판매가 완료되었다”고 언급했다. 이 말은 올해 말 Capa를 기준으로 내년에 생산 가능한 물량이 판매 완료되었다는 뜻이다.

보다 자세히 설명하자면, 장비를 지금 투자하면 리드타임을 반영하여 올해 말 Capa에 영향을 준다. 때문에 현재 확정된 투자 계획을 반영하면 내년 생산이 대부분 예상 가능하고, 이러한 물량이 판매되었다는 의미가 된다. 고객이 만약에 생산량을 넘어 추가로 주문을 한다면, 지금부터 늦게는 올해 하반기까지는 추가 투자가 이루어져야 한다.

**엔비디아 사례로 살펴보는 HBM 수급.** 작년 SK하이닉스의 생산량인 15억 개 Gb가 모두 엔비디아에 판매되었고, 올해 60억 개로 예상되는 생산량 중의 60-70%인 40억 개 수준이 엔비디아에 판매될 것으로 예상된다. 전년 15억 개 대비 2.6배의 성장이다. 엔비디아 판매 물량을 제외한 나머지 20억 개는 AMD와 브로드컴 등의 고객에게 제공된다.

내년에는 약 120억 개 수준의 생산량 중 60-70%가 엔비디아로 할당될 것으로 예상되는데, 이는 80억 개 수준으로 올해보다 2배 성장하는 것이다. 나머지 40억 개가 할당되는 소규모 고객들이 엔비디아를 성장률에서 앞서는 것은 당연하다. 엔비디아의 성장률이 너무 높게 추정된 것이 아니라면 내년도도 공급 과잉은 아닐 수 있다. 물론 다른 공급업체의 과잉 생산이 문제가 될 수 있지만 현재로서는 그 가능성을 높게 보고 있지 않다.

**더블딥은 없다... 2024~2025 사이클은 V자 반등**

대개의 경우 어떤 산업·시장에서 인수 합병으로 업체가 통합(consolidation)되어 그 숫자가 줄면 해당 산업의 기업은 더 높은 마진을 누리게 된다. 작은 수의 업체가 더 많은 물량을 생산하면 각 업체마다 규모의 경제가 커지기 때문에 원가를 낮출 수 있다. 또 수요 상승에 빠르게 대응할 수 있는 잠재적 생산 역량(productive capacity)이 줄어드는 효과를 내기 때문에 가격 방어나 상승이 좀더 쉬워질 수 있다. 주요 플레이어가 삼성전자·SK하이닉스·마이크론의 3개 업체밖에 안되는 디램(DRAM·컴퓨터가 계산할 때 데이터를 일시 저장할 때 주로 쓰임) 시장이 대표적인 예다.

물론 2023년에 디램 업황이 좋지 못했던 것은 이런 효과를 충분히 누리지 못했다고도 볼 수 있다. 그러나 2024년에는 업체들의 자발적인 가동률 조정과 재고 감축에 따라 제품 가격이 상승하고 있다. 가격이 더 떨어지지 않거나 오르기 시작하면, 고객사는 높아지는 가격에도 메모리를 사들여 재고를 쌓게 된다.

**HBM의 Capa 흡수.** 다시 공급이 증가하며 가격이 다시 하락하는 더블딥 우려는 크지 않다고 판단한다. AI용 GPU에 탑재되는 고성능 메모리인 HBM 수요가 급증하고 있는 데다, 이 수요를 대응하려면 CPU용 고성능 메모리 생산까지 제한해야 할 수도 있어 CPU용 메모리 역시 공급 과잉·가격 하락으로 갈 가능성이 낮기 때문이다. 이를 제대로 이해하려면 메모리 제조사마다 첨단 제조 라인으로 만들어낼 수 있는 제품의 수량에 한계가 있고, 특히 HBM을 만들 때 기존 메모리보다 더 많은 생산 역량이 투입되는 것을 알아야 한다.

HBM을 만들 때 기존 메모리보다 더 많은 생산 역량이 투입되는 것은 HBM의 기술적 특성 때문이다. HBM은 빠른 데이터 통신을 위해 디지털 신호의 출입구라고 할 수 있는 'I/O(Input/Output)'를 기존 메모리(DDR)의 8개에서 1,024개로 늘렸다. 도로를 2차로에서 8차로로 넓히면 훨씬 더 많은 차가 같은 시간에 목적지에 도달할 수 있는 것과 같은 원리다. 이를 통해 HBM은 통신 속도는 더 높으면서 발열과 전력 소비를 낮췄다. 더 많은 양의 데이터를 더 빨리 통신해야 하는 AI용 프로세서(예를 들면 GPU)에 HBM이 필수인 이유다.

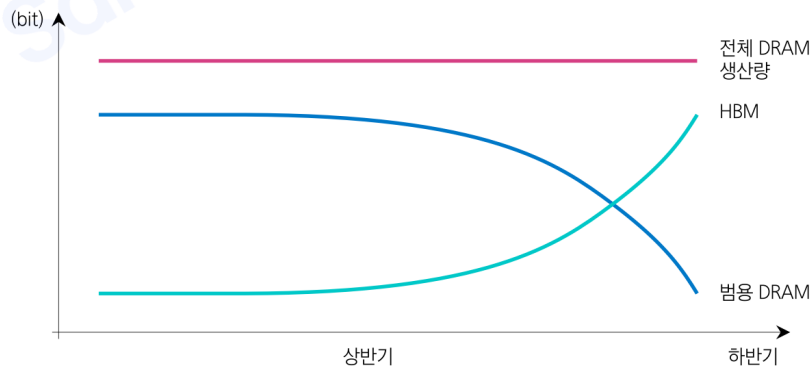
**HBM 1개는 범용 디램 3개 생산 Capa.** 문제는 이렇게 I/O를 늘리는 과정에서 전체 반도체의 크기가 기존 메모리(DDR)보다 배 이상 커진다는 점이다. 메모리 제조 3사 중 한 곳인 미국 마이크론의 첨단 제조공정(1b)을 예로 들어보자. 12인치 웨이퍼(wafer·칩 생산의 기본 재료, 피자라 치면 도우에 해당) 한 장에서 16Gb DDR5를 1,600개 만든다면, 같은 공정에서 24Gb HBM3e는 580개밖에 만들 수 없다고 한다. 같은 기본 재료(웨이퍼)에서 만들 수 있는 칩의 수가 64%나 줄어드는 것이다. 거기다 수율(투입된 제품 수 대비 양품 비율)도 낮다. 대략적으로 1개의 HBM을 만드는 것은 3개 수준의 범용 디램을 만드는 자원이 필요하다.

그런데 HBM을 만들어야 하는 제조 라인은 지금까지 서버용의 고성능 DDR5 모듈(여러 개의 메모리로 구성된 통합부품)을 생산하던 곳이었다. 따라서 첨단 제조 라인에서 HBM을 더 많이 생산하게 되면, 라인을 공유해야 하는 CPU용 고성능 메모리의 생산이 제한될 수 있다. 그렇다면 공급이 수요를 못 따라가는 HBM은 물론, 기존 메모리도 고성능 제품을 중심으로 가격방어가 가능하다는 논리다. 이런 이유로 2024년 가동률이 다시 상승해도 업계의 생산 증가율(%)은 한 자리 수에 그칠 전망이다. 이는 지속적인 재고 소진과 업황의 개선을 의미한다.

**업계의 목표는 최대한 많은 HBM 생산.** 또한 업계는 최대한 모자라는 HBM 생산을 위해 전력을 다하고 있다. 이는 90년대 말 고대역폭 메모리인 Rambus DRAM이 삼성전자와 NEC를 제외한 업체들에게는 환영받지 못했던 것과 대조된다. Rambus는 메모리 공급사가 Rambus에게 로열티를 지불하는 방식이었고, 매년 기술이 업데이트되며 라이선스 계약을 갱신할 경우 새로운 비용이 발생하는 속칭 '노예 계약'이 될 수 있었기 때문이다.

모두가 만들기 어려운 HBM을 최대한 확보하려 한다면 그만큼 기존 범용 디램의 생산은 줄어들게 된다. 당사가 하반기 레거시 디램이 공급 부족을 겪고, 내년에 데이터센터의 증설이나 서버 교체 주기가 도래한다면 공급 부족이 더욱 커질 수 있다고 예상하는 이유다. 지금이라도 공장을 지어 증설을 한다면 대응이 될 수 있으나 메모리 3사 중 땅(장비를 설치할 클린룸 공간)이 여유가 있는 업체는 삼성전자뿐이다.

#### 2024 DRAM 생산량



자료: 삼성증권 추정

## 삼성전자에 대한 Risk Return이 좋아졌다는 판단의 배경

## 하반기로 갈수록 빨라지는 수익성 개선

우리는 HBM3 및 HBM3e 신제품 확대, HBM Capa 증설로 인한 레거시 디램 공급 부족으로 인한 수익성 개선으로 2024년 삼성전자와 SK하이닉스의 연간 영업이익을 각각 43조원(OPM 14%), 22조원(OPM 33%)으로 예상한다. 이는 올해를 기준으로 기존 예상 각각 41조원, 20조원에서 상향 조정된 것이고, 상향 조정의 배경은 하반기 가격 예상이 더욱 긍정적으로 변하였기 때문이다. 우리는 하반기 메모리 가격 상승 가속화에 따라 삼성전자와 SK하이닉스의 연간 영업이익이 컨센서스를 각각 11%, 16% 상회할 것으로 전망한다.

**디램 가격 전망 변화.** 삼성전자 기준, 2024년 디램 가격 전망은 기존 1분기부터 매분기 20%, 15%, 5%, 3% 상승(전분기 대비)에서, 새로운 예상은 20%, 15%, 8%, 10% 상승으로 전망한다. SK하이닉스 기준으로 기존 2024년 디램 가격 전망은 1분기부터 매분기 21%, 15%, 5%, 7% 상승(전분기 대비)에서, 새로운 예상은 21%, 14%, 15%, 15%이다.

**낸드도 동반 상승.** 낸드도 공급 확대로 인한 공급 과잉 우려보단 AI 수요 강세에 대한 확신이 짙어지며 메모리 제조사의 이익 전망치는 지속적으로 상향 조정되고 있다. AI의 온기가 디램뿐만 아니라 낸드로도 확산되고 있다는 점도 메모리 공급사의 수익성 개선에 기여할 것으로 기대된다.

## 삼성전자, SK하이닉스 연간 실적 추정

(조원)	New		Old		컨센서스		차이 (% , 기존 전망 대비)		차이 (% , 컨센서스 대비)	
	2024E	2025E	2024E	2025E	2024E	2025E	2024E	2025E	2024E	2025E
<b>삼성전자</b>										
매출액	308.1	343.4	305.9	336.0	309	342	0.7	2.2	(0.2)	0.5
영업이익	43.2	55.7	41.0	50.1	39	55	5.4	11.0	10.9	1.9
<b>SK하이닉스</b>										
매출액	68.0	91.6	66.5	80.8	64.9	80.6	2.3	13.3	4.8	13.7
영업이익	22.4	31.3	20.2	25.7	19.3	25.9	10.7	21.8	16.1	20.9

자료: 각 사, Fnguide, 삼성증권 추정

## 삼성전자 HBM이 하반기에 고객 인증 통과할 가능성이 있다는 판단

SK하이닉스의 하반기 가격 상승이 삼성전자보다 더욱 가파른 것은 신규 HBM3e의 믹스가 더욱 크고 다른 사업부가 존재하는 삼성전자와 달리 반도체만 있기 때문이다. 하지만 삼성전자도 하반기에는 HBM3e에 대한 추적이 시작될 것으로 예상된다. 보도 등에서 언급된 바와 같이 발열과 전력 소비 등 개선할 점이 있다는 평가도 있지만 이는 상대적인 비교이고, 절대적인 기준에서는 예정된 기한 내로 고객 인증을 통과할 가능성이 있다는 판단이다.

**보다 근본적인 변화도 진행 예상.** 전하의 이동을 빠르게 한다는 것은 인공지능 반도체에서 가장 중요한 것이다. HBM에서는 이를 I/O를 늘려 대역폭을 크게 확대하였다. 하지만 동시에 HBM을 구성하는 각 메모리 layer들에 대해서도 성능을 높여야 한다. 우리가 마당에 물을 주는 파이프에 높은 압력을 가하면 더욱 빠른 물의 이동을 만들 수 있지만, 이는 저항을 높이고 동시에 발열과 전력 소모의 배경이 되기도 한다. 상대적인 의미에서 이러한 점은 더욱 중요하게 다루어져야 한다.

삼성전자는 대표적인 메모리 제조업체로 제조 효율에서 기술적인 선도를 해온 기업이다. 그동안 설비 투자 대비 효율이 낮고 빠르게 변하는 HBM 시장에서 이제는 ROE와 ROIC의 개선을 위한 제조의 효율 개선을 위해서 기존의 속도-발열-전력 소비의 트레이드오프를 원가 혁신에서 속도 개선에 방점을 두는 것이 필요할 수 있다. 이러한 변화는 투자를 집행하는 기준에도 적용될 수 있다.



**무엇이 문제이고, 언제 좋아지는가.** 우리가 반도체 주식을 바닥에서 사는 논리는 더 이상 나빠지지 어렵다는 소위 '2<sup>nd</sup> derivative'적인 접근이다. 오랜 기간 반도체 사이클에서 재고가 고점을 찍으면 좋아질 일만 남았고, 재고 소진이 12개월 이상이 걸려도 저점에서 사는 것이 최선이라는 판단이다. 보수적으로 보면 기술에 대한 이야기는 다를 수도 있다. 기술을 개발하는 사람은 잘 변하지 않고, 따라가기도 어려울 수 있다. 특히 HBM은 고객의 시스템과 공급사의 메모리를 서로 동기화하는 작업이 필요하다. 선도 업체가 아니라면 고객의 테스트 보드가 없어 납품 전 제대로 된 품질 관리가 어려울 수도 있다.

아마 문제가 많기로는 중국 반도체 업체보다 많은 곳은 없을 것이다. 중국 반도체는 반도체 육성 초기 (~2020년)와 현재 많은 차이가 있다고 판단된다. 더욱 중앙정부와 화웨이 등 빠른 의사 결정이 자리를 잡아 반도체 사업을 수행하기 좋아진 것처럼 보이지만, 정작 노하우를 가진 외국인 인력은 더욱 배척되고 있다. 그러다보니, 문제를 해결하는 과정에서 문제에 대한 판단도 해결도 납득하기 어려운 점이 이전보다 오히려 늘어나고 있다는 판단이다. 빈 수레가 요란하듯이 살아남은 중국 반도체 사람들은 모든 것을 할 수 있다는 구호를 되풀이하고 있기 때문이다.

투자자들이 재고 사이클을 두고 12개월 먼저 저점 투자에 나서는 것과 같이, 문제를 파악하고 개선하기 시작한다면 기술에 대한 문제도 크게 다를 수 없다. 재고 사이클도 시스템적인 Risk가 발생하지 않는다는 전제가 필요하고, 기술에 대한 문제도 해당 기업에 같은 문제가 발생하지 않는다는 전제가 필요하다. 여기에 대한 판단은 투자자의 몫이다. 앞으로 6개월 내 많은 추측성 보도와 리포트에 희비가 엇갈릴 수도 있다. 하지만, 문제를 개선할 수 있다는 것이 당사가 지난 오랫동안 삼성전자의 반도체를 두고 관찰한 결과이다. 메모리 셀의 속도, 공정, 품질 관리는 삼성전자가 오랫동안 잘 해오던 것이다.

**투자자들의 선입견과 Risk-Return. 1년 목표가 11만원에서 12만원으로 상향 조정.** 적어도 지난 1~2주 동안 투자자들과 대화를 해보면 투자자는 크게 두 가지 결론 또는 질문을 하고 있다. 하나는 삼성전자의 catch up이 적어도 1년 안에는 쉽지 않다는 생각과, 또 하나는 삼성전자가 경영진 변화로 인해 과거 점유율 싸움을 다시 할 수 있다는 생각이다. 그리고 그 생각은 현재 주가에 반영되고 있다. 우리는 두 가지 생각이 틀릴 가능성이 크다는 판단이다. 앞서 설명한 바와 같이 이익 전망치는 더욱 상향 가능성이 크다. 그리고 기술은 격차가 좁혀질 것으로 생각된다.

2023년 소니가 애플에 이미지 센서를 납품하며 벌어진 사례와 비교해보자. 소니는 센서의 구조를 2레이어에서 3레이어로 변경하며 극심한 수율 부진이 발생하였다. 소니는 이에 애플을 제외한 모든 고객의 주문을 뒤로하고 아이폰 15의 출시에 필요한 납기와 물량을 달성하였다. 큰 비용을 들였지만 고객의 신뢰를 지켜낸 것이다. 삼성전자의 HBM은 경우가 조금 다르다. 경쟁도 치열하고, 동일 제품에 대한 고객도 많다. 때문에 삼성전자는 소니와 같은 극단적인 대처를 하기보다는 HBM은 최대한 품질을 개선하고 물량을 늘려갈 수 있도록 하고, 대신 이로써 발생하는 기회 비용은 범용 메모리 가격의 상승으로 만회할 수 있다는 판단이다.

많은 고객에게 reference를 받을수록 품질 개선은 빨라진다. 현재 삼성전자의 HBM을 둘러싼 많은 부정적인 보도가 있지만, 당사는 오히려 한발짝 물러나 보면 삼성전자에 대한 Risk 대비 Return은 지난 1년 동안 가장 긍정적이라는 판단이다. 이익 전망치를 상향 조정하고, 이를 배경으로 12개월 목표가를 11만원에서 12만원으로 상향 조정한다.

## 삼성전자 실적 전망

(십억원)	1Q24	2Q24E	3Q24E	4Q24E	1Q25E	2Q25E	3Q25E	4Q25E	2023	2024E	2025E
매출액	71,916	73,801	80,287	82,123	79,612	81,383	92,037	90,346	258,935	308,127	343,378
반도체	23,308	27,075	29,374	32,382	31,417	33,612	37,974	37,574	68,467	112,303	140,577
DRAM	10,658	12,918	13,869	15,409	14,797	17,091	20,085	20,457	27,867	52,854	72,430
NAND	7,000	8,300	9,310	10,633	9,908	10,069	11,064	11,141	18,142	35,243	42,182
LSI/파운드리	5,650	5,857	6,195	6,339	6,712	6,452	6,825	5,976	22,469	24,042	25,965
디스플레이	5,390	5,913	8,600	9,834	5,886	5,903	8,944	10,193	30,975	29,736	30,926
OLED	5,390	5,913	8,600	9,834	5,886	5,903	8,944	10,193	30,975	29,736	30,926
IM 사업부문	33,530	29,581	30,881	27,007	30,912	29,466	31,615	27,383	112,410	120,999	119,376
핸드셋	31,517	27,318	28,698	24,824	28,613	27,188	29,359	25,128	103,935	112,356	110,289
CE/Harman	16,680	18,232	18,432	19,900	17,397	18,403	19,504	21,196	70,829	73,245	76,499
영업이익	6,606	8,660	12,600	15,382	12,213	12,832	15,755	15,300	6,567	43,248	55,661
반도체	1,916	4,184	6,816	9,898	7,736	8,698	10,283	9,543	-14,880	22,814	36,261
DRAM	1,969	3,230	4,578	6,473	5,476	7,008	8,839	8,389	-1,132	16,251	29,712
NAND	856	1,458	2,095	3,130	2,132	1,471	1,287	966	-11,529	7,539	5,856
LSI/파운드리	-909	-504	142	295	128	219	157	188	(2,219)	(976)	693
디스플레이	340	444	1,853	1,999	460	472	1,882	2,072	5,566	4,636	4,887
OLED	340	444	1,853	1,999	460	472	1,882	2,072	5,566	4,636	4,887
IM 사업부문	4,070	3,118	2,962	2,443	3,173	2,707	2,453	2,485	13,010	12,593	10,817
핸드셋	3,932	2,955	2,803	2,284	2,998	2,533	2,280	2,312	12,385	11,974	10,123
CE/Harman	770	964	1,020	1,091	894	1,004	1,187	1,250	2,434	3,844	4,336
영업이익률 (%)	9.2	11.7	15.7	18.7	15.3	15.8	17.1	16.9	2.5	14.0	16.2
반도체	8.2	15.5	23.2	30.6	24.6	25.9	27.1	25.4	-21.7	20.3	25.8
DRAM	18.5	25.0	33.0	42.0	37.0	41.0	44.0	41.0	-4.1	30.7	41.0
NAND	11.6	17.6	22.6	29.6	21.6	14.6	11.6	8.6	-65.8	21.4	13.9
LSI/파운드리	-16.1	-8.6	2.3	4.7	1.9	3.4	2.3	3.2	-9.9	-4.1	2.7
디스플레이	6.3	7.5	21.5	20.3	7.8	8.0	21.0	20.3	18.0	15.6	15.8
OLED	6.3	7.5	21.5	20.3	7.8	8.0	21.0	20.3	18.0	15.6	15.8
IM 사업부문	12.1	10.5	9.6	9.0	10.3	9.2	7.8	9.1	11.6	10.4	9.1
핸드셋	12.5	10.8	9.8	9.2	10.5	9.3	7.8	9.2	11.9	10.7	9.2
CE/Harman	4.6	5.3	5.5	5.5	5.1	5.5	6.1	5.9	3.4	5.2	5.7

자료: 삼성전자, 삼성증권 추정

## SK하이닉스 실적 전망

(조원)	1Q24	2Q24E	3Q24E	4Q24E	1Q25E	2Q25E	3Q25E	4Q25E	2023	2024E	2025E
매출액	12.4	16.2	17.9	21.5	20.4	22.2	24.8	24.2	33	68	92
DRAM	7.3	10.1	11.4	14.2	13.5	15.2	17.0	16.5	21	43	62
NAND 및 기타	4.8	5.7	6.1	6.9	6.5	6.7	7.3	7.3	11	23	28
영업이익	2.9	4.9	5.9	8.7	7.2	7.7	8.9	7.4	-8	22	31
DRAM	2.5	4.0	4.9	7.0	6.0	7.0	8.2	7.1	1	18	28
NAND 및 기타	0.4	0.9	1.0	1.8	1.1	0.7	0.7	0.3	-8	4	3
영업이익률 (%)	23.2	30.3	32.8	40.5	35.2	34.8	36.0	30.8	-23.6	32.9	34.1
DRAM	34.4	40.0	42.5	49.0	44.7	46.4	48.3	43.2	2.8	42.7	45.7
NAND 및 기타	8.0	15.3	16.7	25.4	17.5	10.6	9.4	4.3	-78.2	17.2	10.3

자료: SK하이닉스, 삼성증권 추정



## 반도체 소부장 투자 전략

### 하반기 소부장에서 주목할 부분

반도체 호황은 소부장 업계에 긍정적이고, 당사는 하반기 특히 주목할 부분이 다음과 같다는 판단이다.

#### 1) 반도체 장비

- **후공정 확대:** AI 수요에 대응하기 위해 후공정 투자가 늘어나고 있고, 세부적으로 기술 (CoWoS-L/S 적용, HBM I/O 확대 등)에도 변화가 나타나고 있다. 각각 파크시스템스와 피에스케이홀딩스가 수혜다.
- **로직/파운드리 전공정 투자 회복:** 2024년 상반기, 파운드리 업계의 전공정 투자는 매우 부진하다. 하지만, 하반기에는 전공정 투자 회복이 예상된다. 3nm 수요가 제품 및 고객 다변화 효과 속 확대되고 있고, 4/5nm의 가동률은 AI GPU 생산 확대 효과 속 어느덧 추가 증설 필요 구간으로 진입하고 있다.

2) **반도체 소재/부품:** 단기 이익 성장세가 기대 대비 약하다는 점은 아쉬우나, 이익의 방향에 있어 변한 것은 없다. 가동률과 최선단공정 중심 전공정 설비 투자의 회복은 소재/부품 수요 성장을 이끌 요소이고, 낮아진 밸류에이션은 Long Term 투자자에게 좋은 Entry Point로 작용할 것이라 생각한다. 단, 긍정의 강도는 반도체 후공정을 선호한다.

#### Alpha 찾기

반도체 업황이 개선될 때 반도체 소부장은 대부분 같은 방향성을 가진다. 하지만 추가적인 Alpha를 찾는다면, 그간 상대적으로 관심을 덜 받아왔던 삼성전자의 공급망을 주목할 필요가 있다.

- **HBM Value Chain:** SK하이닉스 대비 삼성전자 Value Chain의 약세가 이어지고 있다. 향후 나타날 HBM 투자는 삼성전자 중심으로 나타날 가능성이 크다. 삼성전자 HBM 경쟁력에 대한 전망이 과거 대비 긍정적으로 변하고 있는 만큼 삼성전자 Value Chain의 De-Rating 또한 해소될 것으로 예상된다.

#### 1) 피에스케이홀딩스 (031980 KS/Not Rated): HBM Play의 정석

- 주력 고객사의 대규모 HBM Capa 증설이 눈앞으로 다가온 만큼 단기적으로 이익 모멘텀이 부상할 것이다. 현재 HBM 3사 모두로 장비(Reflow, Descum)를 공급하고 있으며, 국내외 HBM 고객 모두에게 Exposure가 있다는 점은 다른 소부장 주식과 차별화되는 장점이라 생각한다.
- Descum 장비: TSV 공정 후 남은 찌꺼기를 제거하는 데 활용되는 장비이며, 중장기적으로도 물량 성장을 기대해볼 수 있는 라인업이라 생각한다. HBM의 경우, 보다 빠른 속도를 구현하기 위해 I/O 수가 지속 확대(HBM3e: 1,024개, HBM4: 2,048개)될 것으로 예상되고, 늘어나는 I/O의 수는 Descum 장비의 구조적 스텝 성장을 이끌 매개체로 작용할 것이라는 판단이다.

#### 2) 파크시스템스 (140860 KS/목표주가 240,000원)

- 목표주가를 220,000원에서 240,000원으로 상향한다 (Target EPS: 2024년 25%, 2025년 75% 반영). 현 주가 기준 상승 여력은 37%이다. Advanced Packaging (CoWoS)에 대한 기여 확대, EUV 마스크 리페어의 추가 성장 등을 감안 시, 추가적인 주가 상승이 이뤄질 것이라는 판단이다.
- **Advanced Packaging (CoWoS) 기여:** 기존에는 Bonding Thickness 계측 목적으로 활용될 것으로 예상했다면, 인터포저와 HBM을 연결하는 Via Hole의 수직 정렬도 계측 영역에서 활용성이 추가 확대될 것으로 예상된다.
- **EUV 마스크 리페어 장비:** 로직/파운드리 내 고객군이 다변화되고 있고, 추가 오더에 대한 논의가 구체화되고 있는 것으로 추정된다.

## 파크시스템스: 목표주가 산정

구분	비고	
주당 순이익	7,204	2024년 EPS 25%, 2025년 EPS 75% 적용
Target P/E	33.5	2021-2024년 평균 P/E에 1SD를 더한 수준
<b>목표주가</b>	<b>240,000</b>	
현재 주가	174,800	
상승 여력 (%)	37.3	

자료: 삼성증권 추정

## 파크시스템스: 실적 추이 및 전망

(십억원)	1Q23	2Q23	3Q23	4Q23	1Q24P	2Q24E	3Q24E	4Q24E	2022	2023	2024E	2025E
매출액	27.5	39.4	32.4	45.5	25.7	40.1	50.5	57.8	124.5	144.8	174.0	212.2
AFM	14.5	31.5	28.1	35.3	12.4	35.0	39.4	44.3	107.2	109.5	131.0	151.6
연구용	8.4	10.4	10.0	11.1	7.2	12.3	12.6	13.8	37.6	40.0	45.8	47.7
산업용	6.1	21.1	18.1	24.2	5.2	22.7	26.8	30.6	69.6	69.6	85.2	103.9
바이오	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
신규 장비*	9.0	3.0	0.0	3.0	9.0	0.0	5.5	6.0	9.0	15.0	20.5	35.0
기타	4.0	4.9	4.3	7.1	4.3	5.1	5.6	7.5	8.3	20.3	22.5	25.6
<b>영업이익</b>	<b>2.2</b>	<b>10.1</b>	<b>5.5</b>	<b>9.8</b>	<b>0.5</b>	<b>8.1</b>	<b>13.8</b>	<b>16.4</b>	<b>32.6</b>	<b>27.6</b>	<b>38.8</b>	<b>62.7</b>
세전이익	2.7	9.3	6.5	8.7	0.5	8.3	14.5	16.9	31.9	27.1	40.1	64.9
순이익	2.6	7.6	6.3	8.0	0.5	7.3	12.7	14.8	28.0	24.6	35.3	55.2
<b>이익률 (%)</b>												
영업이익	7.9	25.6	16.9	21.6	2.0	20.2	27.2	28.5	26.2	19.0	22.3	29.5
세전이익	9.8	23.6	20.0	19.1	2.1	20.7	28.6	29.2	25.6	18.7	23.1	30.6
순이익	9.6	19.3	19.5	17.6	1.8	18.2	25.2	25.7	22.5	17.0	20.3	26.0

참고: \* 마스크 리페어 장비와 하이브리드 WLI 장비

자료: 파크시스템스, 삼성증권 추정

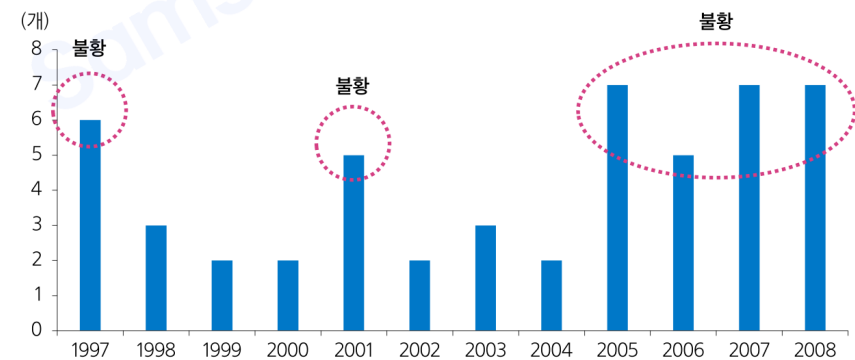
**생산을 늘리고 싶어도 부족한 클린룸 공간**

**Agility is all you need.** 수요 전망이 틀리는 것은 사실 불가피하다고 생각한다. 특히 초기 인터넷과 AI와 같이 급변하는 기술 환경에서 미래의 수요를 정확하게 전망한다는 것은 거짓일 가능성이 높다. 최근 한 공급회사는 아예 2025년 HBM 메모리 수요 전망을 하지 않는다고 했다. 부정확한 가정에 근거해서 경영전략을 세우는 것은 잘못된 결과를 초래할 가능성이 높다는 이유다. 잘못된 정보에 기반한 시스템이 에러를 유발한다는 의미다. 때문에 '기민함 (Agility)'이 더욱 강조된다.

시장은 HBM의 호조로 SK하이닉스가 설비 투자를 급격히 늘릴 수 있다고 보고 있지만, 당사는 최근 SK하이닉스의 2024년 설비 투자가 기존 목표에서 다시 줄어드는 방향으로 재조정되고 있다고 판단한다. 비용과 환율이 오르기 때문이다. 아마도 올해의 매출 대비 설비 투자는 30%를 넘기지 않을 것으로 예상되고, 이는 아주 건강한 향후 수급을 제시한다. 과거 한국, 일본, 대만이 치열하게 경쟁하던 시점과는 확연히 달라진 모습이다.

과거의 증설 경쟁은 팍을 더 짓는 것이다. 사실 반도체 공급은 땅을 확보하고 그 위에 건물을 세우고 클린룸을 만들어 장비를 돌리는 과정을 거치게 된다. 90년대부터 메모리 업체의 사이클은 몇 번의 과잉 경쟁이 동반되었다. 일반적으로 매년 2~3개의 신규 팍을 지었지만 1997년 대만의 진입으로 신규팍이 6개로 늘어났고, 인터넷 열풍으로 2001년 신규 오픈한 팍은 역시 5개에 달했다. 이러한 과정은 이후에도 이어졌다. 휴대폰과 노트북 등 모바일 수요의 확대와 대만과의 경쟁으로 2005년 7개, 2006년 5개, 2007년 7개, 2008년 7개의 공장이 가동을 시작하여 예외 없이 불황으로 이어졌다.

사실 메모리 업계가 3개로 줄어든 이후 신규 팍을 가동하는 것은 상당히 드문 일이 되었다. 막대한 투자 비용으로 대부분의 증설은 공정 전환으로 이루어진다. 올해 초 한국 메모리 업계의 생산량이 전년 대비 60% 수준 늘었다고 하지만, 이는 작년 쌓여있던 재공 재고(웨이퍼 상태)를 만들어 팔았기 때문이고, Capa는 늘어나지 않았다.

**연간 신규 반도체 공장 Fab 가동**

자료: Gartner

**유가 사례에서 보는 시사점: 공급의 경직성과 영향**

공급의 경직성을 완화한다면 (무조건 점유율을 위해 증설하지 않는다면) 향후 한국의 반도체 사업이 더 많은 현금을 얻을 수도 있다. 유가의 사례에서 유사점을 살펴보자. 1990년대 배럴당 10달러 수준이었던 유가가 어떻게 현재 80달러 수준으로 상승했는지에 대한 이야기이다.

유가 사례를 드는 이유는 지정학(地政學, 인문 지리학을 활용해 국제 정치를 분석하는 학문) 혹은 지경학(地經學, 인문 지리학을 활용해 국제 경제를 분석하는 학문)적인 변화가 유가 상승을 이끌었던 과정이 앞으로 메모리 반도체 시장에서 재현될 수 있다고 생각하기 때문이다.

유가는 2000년 이후 브릭스 경제 성장이 수요증가를 이끌고 오펙(OPEC·석유 수출국 기구) 중심으로 공급이 제한되며, 1990년대 배럴당 10달러 수준에서 현재 80달러 수준으로 상승했다. 원유 사용량이 늘어나고 생산기술 혁신에 따른 원가가 하락하며, 사우디아라비아와 같은 산유국은 막대한 부를 축적했다.

**공급이 제한되면 벌어지는 효과.** 유가 사례에서 얻을 수 있는 시사점은 '수요는 증가한 반면, 공급은 제한됐다.'는 것이다. 디램 산업도 마찬가지로 상황이 벌어질 가능성이 있다. 장기적으로는 특히 AI 개발·운용과 관련 디바이스에 필요한 고부가가치 제품 중심으로 꾸준히 늘어날 것이다. 문제는 공급 쪽이다.

만약 자발적 감산과 과잉 투자 자제가 이어지면, 메모리 산업의 현금 창출 능력이 과거 고점을 넘어 크게 상승할 가능성도 있다.

**석유와 반도체, 공통점은 수입.** 또 하나 중요한 것은 석유와 반도체 모두 대부분의 국가가 수입한다는 점이다. 개별 국가는 물론, 특정 지역권에서 자급자족할 수 있는 것이 아니다. 모두가 반도체를 만들겠다고 하지만 특히 반도체 독자 제조를 실현하기 어렵다. 갈수록 고성능과 높은 제조 비용이 요구될 뿐 아니라, 제조에 필요한 인력과 시설, 공급망 등을 단기간에 구축하는 것이 거의 불가능하기 때문이다. 파운드리와 메모리 분야에서는 이미 수십년간 경쟁력을 높여온 대만과 한국의 지위를 넘보기 어려워 보인다.

다시 정리하면 메모리 산업이 공급의 경직성을 완화할 수만 있다면, 지금까지 경험하지 못한 현금을 창출할 가능성이 있다. 산유국이 경험한 석유 가격 상승으로 인한 구조적인 부의 축적을 메모리 산업에서 재현하는 것도 꿈이 아닐 수 있다. 또한 AI 사이클에 민감한 반도체 산업이 변동성을 줄여갈 배경이 된다.

**정치와 경제적 변화가 유가에 미친 영향**

자료: Markets minute, 삼성증권

## Sizing AI. 인공지능 혁신은 상상 이상일 것

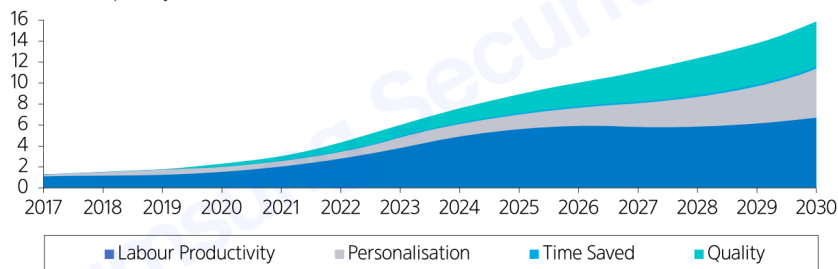
## 인공지능의 가치를 돈으로 계산하면 “중국 GDP만큼”의 게임 체인저

AI의 가치는 40% 생산성과 60% 서비스. PwC(Pricewaterhouse Coopers)컨설팅은 인공지능(AI)이 경제(글로벌 GDP)에 기여하는 가치가 2030년 \$15.7조에 이를 것이라고 전망하였다. 이는 현재 중국의 전체 GDP \$18조와 유사한 규모다. 지난 20년 전 세계 GDP가 미국을 중심으로 한 IT의 혁신을 배경으로 이를 제조하는 브릭스 (브라질, 러시아, 인도, 중국) 특히 중국의 성장이 눈부셨다면, 앞으로는 인공지능을 통해 2030년이 되기 전에 또 하나의 중국이 생겨난다는 예상이다. 예상의 근거를 제공하기 위해 PwC에서는 발생하는 \$15.6조 중 42%인 \$6.6조는 생산성 혁신에서, 58%인 \$9.1조는 새로운 소비를 통해서 발생할 것으로 예상하였다.

또한 이러한 인공지능으로 인한 성장의 효과는 인공지능에서 가장 앞서고 있는 미국과 인공지능에 적용할 대상이 많은 중국에서 70%의 경제적 효과가 집중될 것으로 전망하고 있다. 각국의 경제에 대한 효과로 따지면 미국보다 중국이 더욱 많은 효과를 누릴 것이라는 설명이다. 유럽과 선진국에 포함되는 동아시아는 상대적으로 변화가 작다. 개발도상국은 인공지능에 대한 도입이 상대적으로 느려 가장 경제 성장의 효과가 떨어진다. (p. 67 부록2. 인공지능 밸류체인 참고)

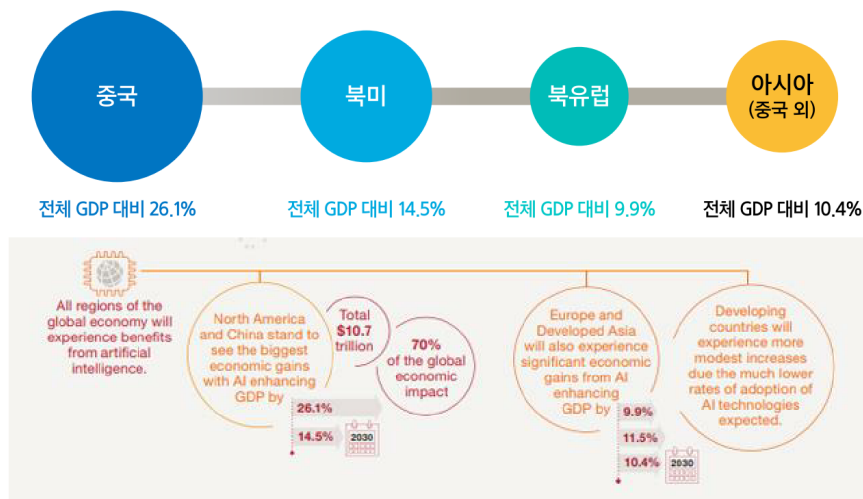
## AI의 가치를 계산한다면 2030년까지 \$15.7조

Global GDP impact by effect of AI (\$조)



자료: PWC

## AI로 인한 지역별 성장 구도



자료: PWC, 삼성증권

**IT의 혁신은 “생산성의 가치”**

생산성 혁신이란 반복된 작업을 자동화하거나, 작업의 효율을 높여주며 대신 보다 높은 부가가치의 작업을 유도하는 것이다. 과거 워드나 엑셀 작업이 사무실의 급격한 생산성 증가를 가져왔듯이, 향후 인공지능이 수반된 컴퓨팅은 사무실에서, 그리고 제조나 유통 현장에서 부가가치를 창출할 것이다.

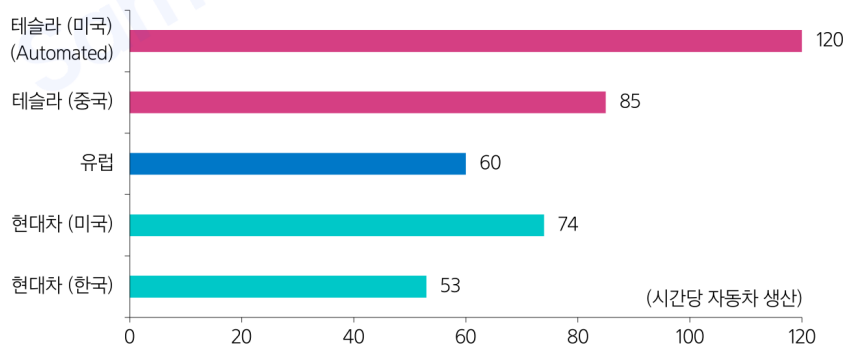
과거 세계 경제 성장의 동력이 구소련이나 중국에서 저가 노동력을 공급하며 생산성의 향상을 제공하였던 것과 같이, 인공지능을 배경으로 자동화된 시스템은 이와 유사한 또는 그 이상의 생산성의 혁신을 약속하고 있다.

**생산 대수의 혁신.** 간단한 예를 들어 보면 시간당 자동차 생산 대수는 과거 지역별로 큰 차이를 보여주어 생산 거점을 생산성이 높은 곳으로 이동하여 왔다. 하지만 이제 인공지능과 함께 이러한 지역별 생산성의 차이는 완전히 다른 양상을 보일 수 있다.

한국에서 현대자동차가 시간당 생산하는 자동차는 53대다. 이는 현대가 미국 공장에서 생산하는 74대보다 낮다. 그만큼 한국은 이제 많은 규제 등 다양한 요인으로 노동 생산성의 경쟁력은 떨어진다. 이는 자동차에만 국한되는 것은 아니다.

유럽의 자동차 업계가 시간당 생산하는 자동차는 60대 수준이다. 현대차의 한국 공장과 비슷하지만, 제조에서 우수한 경쟁력을 바탕으로 성장한 한국 기업의 한국 내 공장이 세계에서 가장 낮은 수준의 생산성을 보이는 것은 놀랍지만 사실이다.

반면, 테슬라가 중국 공장에서 생산하는 자동차는 시간당 85대다. 현대차의 한국 공장보다 60%가 높다. 24시간 3교대로 가동하기에 가능하다. 물론 다른 제품이지만 생산성의 차이는 원가에 큰 차이를 가져온다. 테슬라는 향후 인공지능을 활용한 자동화를 통해 시간당 생산성을 120대까지 달성하겠다고 약속하고 있다. 120대라면 현재 한국보다 생산성이 130% 수준으로 상승하는 것이다.

**인공지능을 활용한 생산성 혁신의 사례: 시간당 자동차 생산량**

자료: 언론 종합, 삼성증권

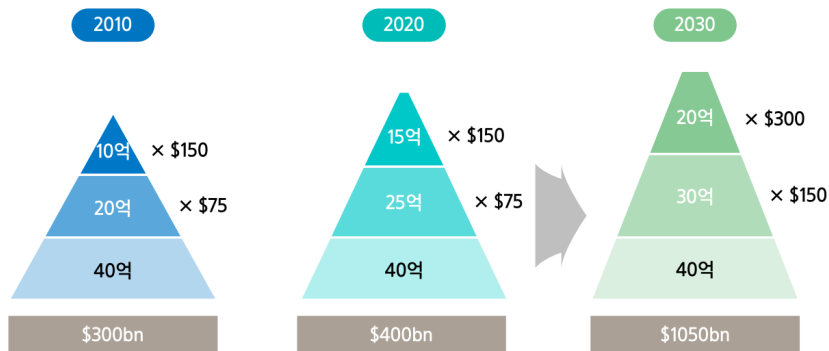
**인공지능이 창조하는 서비스는 “상상한 것 이상일 것”**

생산성 혁신은 비용 대비 뛰어난 효율을 제공하는 새로운 서비스를 가능하게 한다. 새로운 소비를 통하여 발생하는 수요는 소비자가 더욱 양질의 개인화된 서비스를 통하여 발생할 수 있는 새로운 수요다. 예를 들어 테슬라 자율주행을 통하여 운전하는 시간을 아낄 수 있다면 우리는 그 시간에 다른 부가가치를 만들어 낼 수 있다.

반대 의견도 있다. 인공지능이 제공하는 생산성과 이를 기반으로 하는 서비스는 많은 시간이 지난 후에나 가능할 것이라는 우려다. 현재 GPT가 회의 내용을 요약해주거나 그림을 그려주는 것으로 우리의 생활은 크게 달라지지 않는다고 할 수 있다. 1Q24 실적 발표 이후 Meta의 주가가 크게 하락한 것도 Meta가 인공지능을 위한 설비 투자를 기존 \$30~37bn에서 \$35~40으로 확대하였지만, 당장 매출로 인식할 수 있는 서비스는 없다는 것이 배경이다. 특히 Meta AI등 인공지능 플랫폼을 오픈소스로 공개함에 따라 비용 대비 효과를 불신하는 의견이 커졌기 때문이다.

**완전히 뒤바뀌는 시장 구도.** 당장 사용 가능한 서비스가 없다고 기술 개발을 뒤로 할 수는 없다. 당장 사용 가능한 서비스에 집중할 필요도 없다. 90년 말 인터넷의 도입에서 많은 기업들은 인터넷을 통해 모든 것이 이루어진다고 믿었다. 이는 그동안 정보 전달의 매개로 쓰였던 기기들이 PC로 대체됨을 뜻한다. 그러나 당장 인터넷과 PC를 통해 콘텐츠가 전달된 것은 아니다. 미디어 콘텐츠의 유통은 VHS 테이프에서 CD나 DVD로 전환되었고 인터넷을 통한 유통은 고속인터넷과 스마트폰이 정착된 이후다. 하지만 인터넷을 통한 미디어 유통의 혁신보다 새로운 전달 기기의 혁신에 집중했다면 넷플릭스는 나오지 못했을 것이다.

20년이 훌쩍 지나서야 원격 의료와 인공지능을 통해 미리 질병을 센싱하고 오진의 가능성을 낮추고 멀리 떨어진 곳에서도 의료 서비스를 제공할 수 있게 되었다. 하지만 AI가 정작 의료에서 역할을 하는 것은 엄청나게 많은 추론을 통한 신약 개발과 이로 인한 생명 연장이라고 생각한다. AI 혁신을 위한 사업을 구상하는 많은 경우도 “개선”이 아닌 “완전히 다른 형태”로 발전할 가능성이 크다. AI로 인한 인구 피라미드의 변화가 발생하면 관련 산업의 엄청난 성장이 동반된다. 예를 들어 현재 \$500bn 규모의 반도체 산업은 고소득층과 중위소득층이 각 년 \$150, \$75의 반도체를 소비한다고 볼 수 있다. 소비 인구의 수명이 연장되면, IT 소비를 촉진할 수 있다.

**반도체 시장 규모: 인공지능을 통한 상위층의 규모와 소비의 확대**

자료: 삼성증권



**IT 혁신은 “트랜지스터의 혁신”: 연산 컴퓨팅에서 데이터 컴퓨팅으로.**

생산성 혁신의 중심에는 트랜지스터의 혁신이 있다. 과거 계산기에서 PC로, 그리고 스마트폰으로 이어진 혁신도 모두 고성능 연산 기능과 저전력 반도체의 혁신을 통해 가능하게 되었다.

**과거와 차이점은 데이터.** 과거 컴퓨팅을 연산 컴퓨팅이라고 한다면 현재의 컴퓨팅은 데이터 컴퓨팅이다. 과거 인공지능과 현재의 인공지능의 차이점도 활용 가능한 데이터가 많아졌다는 점이다.

PC는 보다 강력한 연산을 위해 클럭 속도를 높이고 더 많은 디램을 사용하였다. 모바일은 저전력이 가미되었지만, 크게 다르지 않다. 이와 달리 AI는 훨씬 많은 데이터의 연산을 필요로 한다. 때문에 GPU와 같은 병렬 컴퓨팅과 이를 지원하는 넓은 대역폭이 더욱 크게 부상하였다.

PC에서는 펜티움이 세대를 더해가며 더욱 빠른 클럭과 연산 속도를 제공하였고, 디램은 이를 지원하기 위해 모듈 단위로 더욱 많은 탑재량을 요구하였다. 과거 데스크탑을 분해하여 성능을 높이기 위해 CPU 옆에 세로로 디램 모듈을 더욱 높은 용량으로 갈아 끼우던 것이 바로 그 배경이다.

모바일에서는 작은 폼팩터를 배경으로 충분한 배터리 사용 시간을 확보하기 위해 저전력 공정이 부각되었다. CPU는 AP로 변신하였고, Operating system은 윈도우에서 가벼운 iOS나 안드로이드로, 디램은 5V 대신에 3V에서 구동하는 모바일 디램으로 전환되었다. 통신이 세대를 더해가며 추가되는 RF 칩들이 소비하는 전력을 줄이기 위해 칩들간 통합이 이루어졌다.

인공지능은 PC와 모바일을 통해 축적된 많은 데이터를 훨씬 효과적으로 연산할 수 있는 GPU와 같은 고대역 반도체의 혁신을 통해 발전하고 있다. 메모리도 High Bandwidth Memory(HBM)으로 기존 32개보다 훨씬 많은 1,024개의 핀을 지원하며 더욱 많은 데이터의 병렬 연산을 지원하고 있다. 그리고 이 같은 트랜지스터의 혁신은 과거 PC의 혁신에서 보여준 무어의 법칙보다 더욱 빠르다.

**데이터 컴퓨팅의 혁신은 "공간의 혁신": 정보화 사회에서 융합의 사회로**

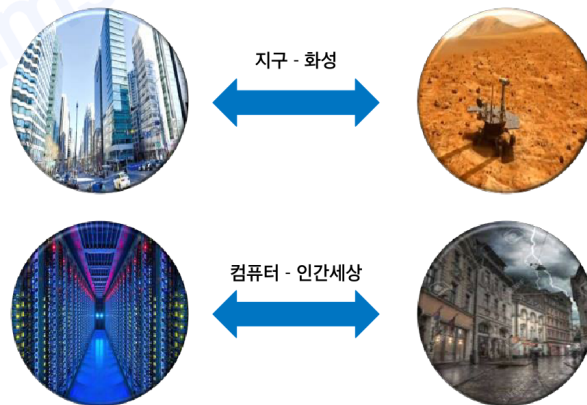
과거 PC와 스마트폰이 정보화 사회를 이끌었다면 미래의 AI가 촉발하는 서비스의 혁신은 융합의 사회를 이끌 것으로 판단한다.

**정보화 사회.** 정보화 사회란 많은 정보를 보다 손쉽게 취득하고 활용하는 것이다. 과거 도서관에서는 주제를 나누고 각 주제별 세분화된 분류에 의해 도서를 분류하였다. 90년대 중반부터는 도서와 신문 등 정보를 컴퓨터에 스캔하고 이를 컴퓨터 시스템으로 살펴보는 시스템이 도입되었고, 이는 구글 서치와 같은 효율적인 정보의 제공으로 이어졌다. 스마트폰이 도입되어 장소에 구애되지 않는 정보의 활용이 이루어졌다.

**데이터의 융합.** 인공지능을 통한 컴퓨팅은 이제 단순히 컴퓨터 안에서 이루어지는 디지털 데이터의 혁신에서, 우리가 살고 있는 물리적인 세계의 데이터를 접목하는 융합의 사회를 의미한다. 데이터의 융합이 이루어진다는 것은 컴퓨터 입장에서 보면 공간의 혁신을 의미한다.

하지만 공간의 혁신에는 많은 어려움이 존재한다. 자동차도 주행 데이터가 필요하고, 로봇도 사람이 움직이는 불특정하고 불확실한 환경에 대한 데이터가 필요하다. 우리가 사진을 찍어도 좋은 사진을 얻기 위해 많은 노이즈를 극복해야 하는 것처럼, 성공적인 AI를 구현하기 위해서는 현실 세계의 많은 노이즈를 극복해야 한다.

완전히 새로운 노이즈의 환경이란 지구에서 활동하는 로봇이 마치 화성에서 완전히 다른 대기와 환경을 경험하는 것과 같은 정도일 것이다. 애플의 음성 인식 비서인 Siri (시리)가 서버에 살다가 우리 현실에 내려오는 순간 주변의 소음과 다른 말투와 억양 등을 극복해야 한다. 그렇지 않다면 AI는 통제되지 않았다고 여겨지거나 많은 실수를 범할 수 있다. 사회적으로 사용 가능하다고 인정되는 합의를 확보하면 우리의 물리적 세상은 크게 변할 것이다.

**컴퓨터 세상 속 인공지능이 인간세상에서 느끼는 낯설음 vs 지구와 화성의 다른 환경**

자료: 구글이미지, 삼성증권

**공간의 혁신을 통한 인공지능은 과거 “PC의 혁신보다 더 강력한” 성장 모델**

PC와 스마트폰, 그리고 인공지능의 차이는 적용되는 공간의 혁신이라고 볼 수 있고, 그 혁신의 결과를 우리가 체감하는 방법은 우리가 인공지능으로 인한 서비스를 사용하는 시간일 것이다. 컴퓨터 등 가상 공간은 우리가 사용하는 시간에만 적용되지만, 현실 세계와 융합된 인공지능은 더 많은 공간을 대체하며 더 촘촘하게 더 넓게 소비자의 지갑을 열 수 있다.

스마트폰은 우리가 사용하는 시간 동안 가치를 창출하지만, 자율 주행은 우리가 차량을 사용하지 않는 시간에도 스스로 돌아다니며 경제활동을 할 수 있다. 스마트폰을 보는 시간을 넘어, 차량과 공장에서도 그리고 가정에서 우리가 어디에 있든지 활용할 수 있게 되는 시스템은 과거의 PC보다 훨씬 강력한 성장 모델이다.

또한 테슬라가 언급한 것처럼 “분산 컴퓨팅”은 각각의 테슬라 차량이 보유한 강력한 추론 컴퓨팅 자원을 연결하여 필요한 사람에게 제공할 수 있다는 것이다. 이러한 개념은 새로운 것이 아니다. 이미 90년대 말에도 연결된 PC의 자원을 공유하여 서버급의 자원을 개인이 확보할 수 있다고 하였다.

사실 지금과 같은 인공지능이 현실화된 것도 일론 머스크와 같은 몇몇 개척자들이 아낌없이 인공지능에 투자하였기 때문이다. 우리는 지금까지 리더십을 보여주는 사람들이 제시하는 방향성을 믿고 기술에 투자하며 발전해왔다. 그리고 현재도 그러고 있기에, 공간의 혁신이 가져올 성장 모델은 근거 없는 허풍이 아니라고 판단한다.

스탠포드에서 발간한 2024 AI 인덱스에 제시된 몇 가지 의미 있는 변화를 보면, 얼마나 많은 자본이 경쟁적으로 투입되고 있고 동시에 기업들이 이를 통하여 매출 성장과 비용 절감을 얼마나 확신하고 있는지 알 수 있다. (다음 페이지 그래프 참고)

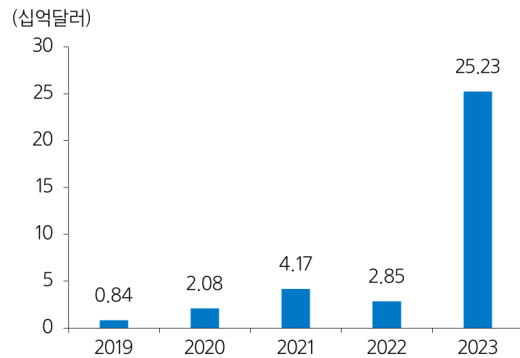
**현실 세계와 융합된 인공지능은 국가의 경계를 넘어 확장.** 인공지능은 우리의 시간을 더 많이 확보하는 것에 그치지 않고, 국가를 넘어 훨씬 넓은 지역의 고객을 커버할 수 있다.

우리가 경험한 경제 성장의 모습은 한 국가가 노동이나 노동의 결과물인 제품을 수출하는 식이었다. 이제 AI의 확대는 사회의 시스템을 수출하는 모습이 될 수 있다. 우리가 일본의 편의점을 한국에서, 한국의 버스와 마트를 몽골에서 볼 수 있듯이, 이제 하나의 도시 기능을 다른 도시에서 낮설지 않게 확인할 수 있다.

한 나라가 다른 나라에 사회의 시스템을 수출한다는 것은 노동력이나 제품보다 훨씬 수명이 긴 부가가치 발생을 초래한다. 마치 제품을 팔아 버는 것이 주식의 가치가 올라 차익을 내는 것이라면, 인공지능을 통한 시스템의 수출은 해당 국가 성장의 지분을 가져가는 채권투자와 같다. 인공지능을 통한 수출과 성장은 보다 지속적이기에 각국은 기술 표준을 내재화하기 위해 치열한 경쟁을 하고 있다. 엔비디아는 “모든 국가들은 소버린 (Sovereign) AI가 필요하다.”고 말했다. 데이터와 인공지능에 대한 국가의 소유권을 잃어버린다면 다른 시스템의 가치에 예측될 가능성이 크기 때문이다.

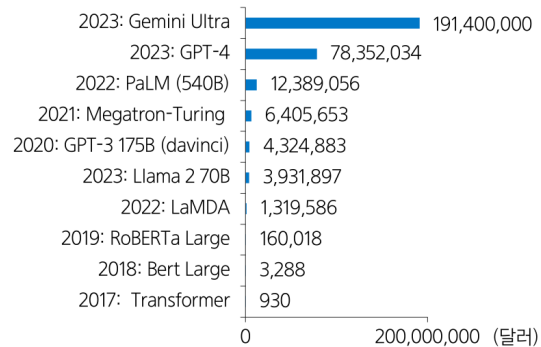
안드로이드는 오픈소스로 공개가 되었지만, 2016년 오라클과의 소송전에서 매출과 손익이 밝혀진 적이 있다. 놀라운 점은 이익률이 무려 70%가 된다는 점이다. 2022년에는 구글이 한국에서만 플레이스토어를 통해 8조 5천억원의 매출을 올렸다는 것도 밝혀졌다. 매년 추가적인 비용은 들어가지 않지만 구글은 모바일 인터넷의 성장을 통해 엄청난 수익을 올리고 있는 것이다.

## AI 민간 투자 규모



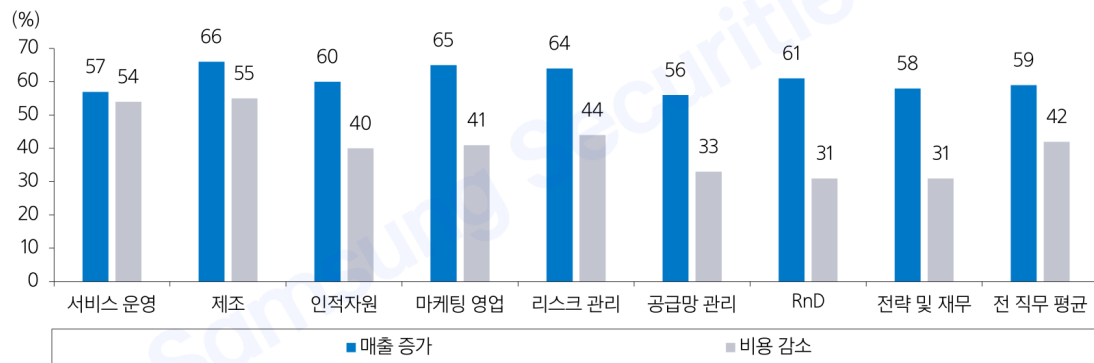
자료: 2024 AI Index report

## 모델별 훈련 비용



자료: 2024 AI Index report

## 업무별 AI 도입 후 매출이 증가 혹은 비용이 감소했다고 응답한 비율



자료: 2024 AI Index report

**AI 시대를 위한 데이터의 통로 확보가 새로운 헤게모니 대결**

우리는 AI 모델과 반도체를 연결하는 컴파일러 플랫폼이 On device AI 헤게모니를 차지하는 것이 가장 가능성이 높다고 생각하지만, 반도체 업체들, 세트 업체들, 그리고 소프트웨어 업체들이 모두 각자의 영역에서 데이터의 통로를 확보하려고 노력하고 있다.

퀄컴은 그 목표에 가장 다가간 업체라고 생각한다. 오랫동안 On device AI에 공을 들이기도 했고, 몇 가지 조건도 갖추었다. 1) PC와 모바일, 자동차, IoT 프로세서를 모두 만든다. 2) 멀티 디바이스의 데이터를 연계한다. 3) 미들웨어에 NPU SDK를 공급한다. 4) 제3자 공급자로, 마이크로소프트, 구글, 오포와 비보 등과의 파트너십이 자유롭다. 반면 OS를 만들지 않는다는 한계도 분명하고, 모바일의 퀄컴에겐 PC에서 인텔과 같은 헤게모니가 없다는 점, 그리고 삼성, 애플, 샤오미, 화웨이 등 유력 모바일 세트 업체들이 수직 통합(vertical integration)을 통해 독자 노선을 시도한다는 것도 위협 요소다.

중국 세트 업체들은 소프트웨어와 서비스의 내재화로 극복하려고 한다. 이들의 목적은 데이터와 서비스의 통로 단일화다. AI 서비스는 당연히 분화될 것이고, 반도체와 OS는 미중 무역 분쟁으로부터 자유롭지 못하다. 결국 그들은 OS를 내재화하여 데이터와 서비스의 플랫폼을 잡는 것이 가장 합리적인 전략이다. 멀티 디바이스에서의 데이터를 잡아야 인공지능 서비스나 메타버스 등 향후 AI 시대의 다음 목표를 세우기도 용이하다. 다만 이들 세트를 이용하지 않거나, 하나라도 다른 브랜드 제품을 사용할 경우 통합 OS의 전략은 크게 희석된다.

**AI의 하드웨어와 소프트웨어와 서비스를 모두 만들 수 있는 업체의 등장**

결국 공은 다시 애플로 돌아간다. 스마트폰에서 소프트웨어와 하드웨어를 동시에 수반하는 기업은 애플이 거의 유일하다. 새로운 기술이 떠오를 때 이러한 강력한 기업이 없다면 기술을 도입하기는 예상보다 더욱 많은 시간이 필요하다. 소위 최적화가 이루어지지 않으면 그만큼 소비자의 선택을 받기 어렵다.

인공지능에서 이 두 가지를 동시에 수행하는 기업은 누구일까?

(1) **OpenAI:** ChatGPT를 선보인 기업이고 가장 가까이 다가와 있다고 판단한다. 디바이스, 로봇, 퍼스널 데이터 관리까지 모두 관심이 있다는 것은 장점이다. 그러나 반도체가 너무 늦었다는 것이 약점이 될 수 있다. 마치 2013년의 테슬라처럼, 디바이스 제조 기술이 검증되지 않았다는 것도 약점이다. 따라서 OpenAI는 TSMC나 삼성전자와 같은 기존의 반도체 제조업체와의 협력을 중요하게 생각하고 있다.

(2) **구글:** AI 시대를 열기에 충분한 요소 기술을 보유했다. 하드웨어와 소프트웨어, 디바이스와 클라우드의 균형을 가진 몇 안되는 업체다. 그러나 인공지능에서는 창의성과 모험심이 부족하여 혁신 기업이 될지 의문이다. 혁신 기업과 경쟁하는 훌륭한 2등 업체가 될 것이 보다 현실적이라는 판단이다.

(3) **테슬라:** OpenAI와 비슷한 지향점을 향하고 있으나 모바일 기기에 대한 연구가 없다. AI 시대를 열기 위해서는 24시간 함께 하는 디바이스를 확보해야 한다. 이 점에서 테슬라는 가장 많은 현실의 물리적인 데이터를 확보한 기업이기도 하고, 또한 분산 컴퓨팅을 통하여 운행을 멈춘 차량이 가용할 수 있는 컴퓨팅 파워를 서로 공유하여 추가적인 매출에 기여한다는 생각이다.

(4) **애플:** 애플의 신제품이 보이지 않게 된 것은 오래되었다. 오히려 새로운 것을 만들 준비가 된 경영진인지 의심이 된다는 것이 시장의 눈초리다. AI 소프트웨어 기술도 뒤쳐진 것으로 판단한다. 그러나 여전히 하드웨어 디자인은 세계 최고이며, AI에 걸맞은 하드웨어와 그리고 무엇보다 중요한 소비자의 감성을 만족하게 하는 소프트웨어라는 정답을 찾아낼 가능성이 있는 회사다. 반도체부터 OS, AI 모델, 서비스까지 수직 계열화할 뿐만 아니라 프로세서의 전성비, 메모리 최적화, 멀티디바이스 환경 등 실사용 측면에서 성능 우위에 있을 것이다. 다만 AI 모델과 서비스 개발이 경쟁사 대비 늦었고, 많은 자원을 투입한 비전 프로의 판매 계획도 비관적이라는 것이 약점이다.

(5) **메타 플랫폼스:** 철저한 하드웨어 파트너십 전략으로 AI에 많은 노력을 기울이고 있으며 킬러 서비스가 존재한다. AI 수혜 기업이 될 가능성은 높지만, AI 시대를 열 수 있는지는 의심된다.

(6) **마이크로소프트:** OpenAI와 소프트웨어/서비스는 앞설 가능성이 있고, 누구보다도 OpenAI의 혁신으로 수혜를 볼 수 있다. 실제 HBM 등 인공지능을 위한 하드웨어 주문 동향에서 엔비디아를 제외하면 가장 적극적인 기업 중 하나가 마이크로소프트다. 그러나 하드웨어 자체에 관심이 없다는 점이 단점이다.

## 단기 Risk: 2000년 IT 버블 vs. 현재

뺏아 있는 2000년과 2024년 시장 분위기. 현재 주식시장이 90년대 말 IT 버블을 뺏아 있다는 의견이 있다. 모두가 경쟁하듯이 인공지능에 대한 투자를 늘리고, 특히 엔비디아 GPU를 비싼 가격에 구입하였다면, 과거 인터넷 버블이 터지고 난 후 공급 과잉이 EMC의 스토리지 시스템, 선 마이크로의 서버, 또는 파이버 옵틱 케이블과 PC의 재고로 이어진 것과 같이, 이번에도 기업들의 인공지능에 대한 FOMO 현상으로 우리는 예상치 못한 환경의 변화 (유가상승, 금리 상승 또는 장기적인 고금리 등)에 따라 갑작스런 기업의 투자 축소가 발생할 수 있고, 이는 GPU와 HBM의 주문감소로 이어질 수 있다.

지난 2022년부터 반도체 업황이 급락한 것이 코로나 기간 동안 PC와 가전제품 등을 실수요보다 많이 구매한 것이 화근이었다면, 이제는 실수요보다 많이 인공지능을 위한 부품과 장비 등을 구매한 것도 화근이 될 수 있다.

기록에서 찾아보는 유사점. 90년대 말 기록에서 지금과 비슷한 몇 가지 정황들을 목격할 수 있다.

우선, 조사기관들은 2000년까지 적어도 2002년까지는 반도체 호황이 이어질 수 있다고 입을 모았다. 이는 대부분의 전망이 공급자 위주로 이루어지기 때문이다.

2000년 7월 TSMC는 ASML에게 첫 12인치 노광장비를 주문하였고, 2002년까지 호황을 장담한 것은 이 장비들이 생산에 적용되기에는 2년의 시간이 필요했기 때문이다. 수요면에서는 인터넷 확대를 위해 PC는 성장하고 인터넷을 배경으로 새로운 Software는 늘어나고, 때마침 모니터도 CRT에서 디지털인 LCD로 전환되고 있었으며, 핸드폰 수요도 늘어나고 있었기에 수요를 걱정하는 사람은 많지 않았다.

## 2000년 당시 기업 탐방 노트 - 호황의 기록들 (1) TSMC: 2002년까지는 문제 없다

□ TSMC

ASML receive 1st order for 300mm from TSMC

○ NEWS: ASML receive 1st order for 300mm from TSMC.

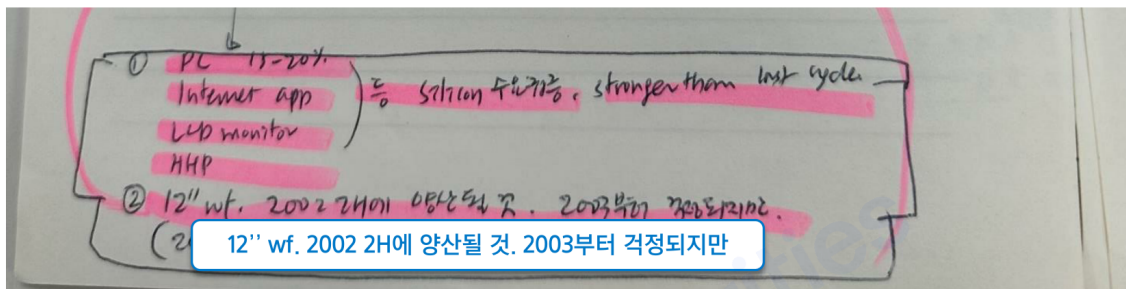
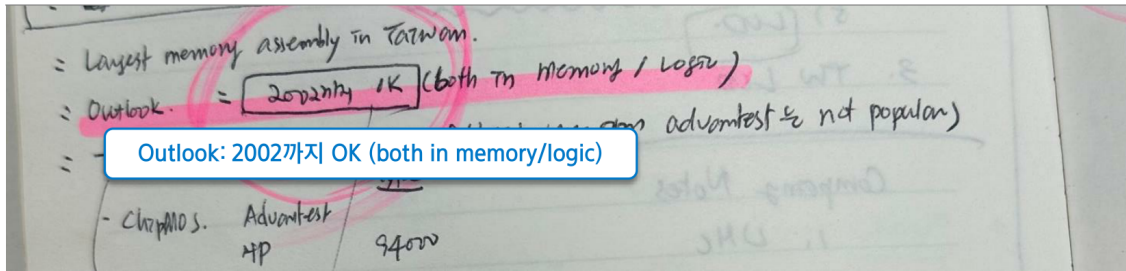
- new pilot line on Tainan (Feb 6)
- 1st shipment scheduled for 100.3Q.
- Capacity = 25/M
- Cost =
- Berzle =
- Process =
- Schedule = 2002 m/p

\* may be after 0.15um

자료: 개인 소장

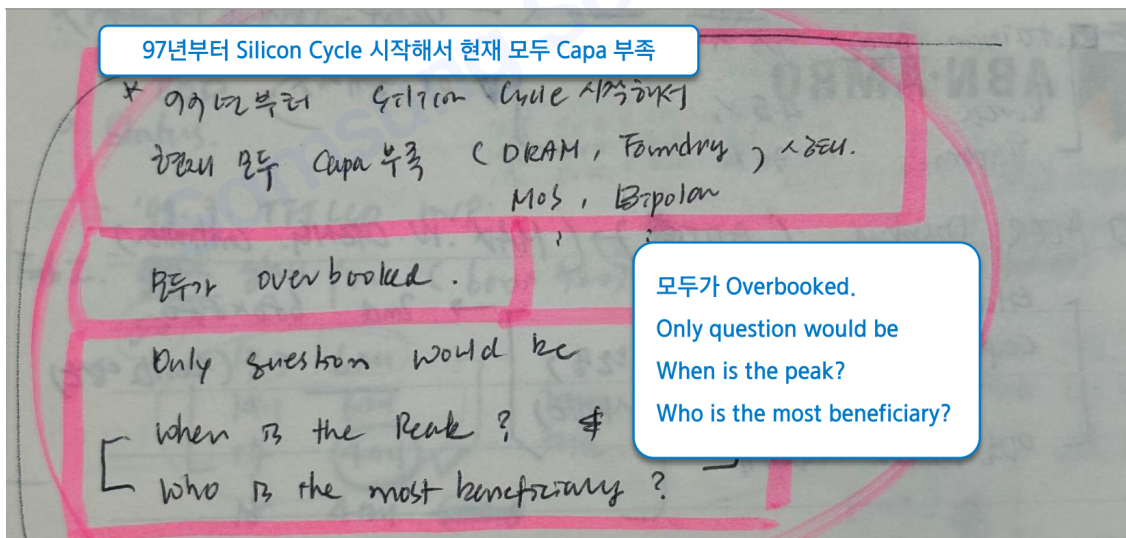


2000년 당시 애널리스트의 개인 노트 - 호황의 기록들 (2) 대만 메모리 업체들: 지난 사이클보다 강하다



자료: 개인 소장

2000년 당시 애널리스트의 개인 노트 - 호황의 기록들 (3) 시장 조사기관: 모든 반도체가 부족하다



자료: 개인 소장



### 사이클의 변화는 민감한 (Extreme) 부분에서 시작

**Rambus와 HBM, 같은 High Bandwidth Memory.** 사이클의 변곡점에 가장 민감하게 반응하고 따라서 사이클의 변화를 가장 먼저 감지할 수 있는 곳은 공급망의 Extreme한 곳, 즉 가장 공급이 부족하거나 많은 제품의 수급에 대한 변화다. 공급망계도 90년대 말과 현재 모두 컴퓨팅에서 우리는 고대역(High Bandwidth)이라는 신제품에 가장 큰 부가가치를 지불하고 있다.

인터넷의 초기와 AI의 초기, 우리는 모두 대역폭을 늘려가야 하는 필요성을 느끼고 있다. 그만큼 컴퓨팅에 많은 데이터를 적용하기 때문이다. 인터넷 초기에는 텍스트에서 그림이나 동영상으로 데이터가 변화하며 고대역이 필요하였고, 현재는 더욱 많은 파라미터의 LLM을 구동하기 위한 고대역이 필요하다.

현재 고대역인 HBM은 과거의 고대역인 Rambus와 다르다고 할 수도 있다. 하지만 두 제품의 초기가격 구조와 원가구조는 아주 비슷하다.

**높은 후공정 비용, 높은 가격, 낮은 생산성에서 모두 닮은 점.** 현재 HBM이 Package 원가가 기존의 5배, HBM의 가격은 약 6~7배 비싸고 전 세계 공급을 90% 이상 한국의 두 업체가 독점하고 있다. 하지만 90년대 Rambus도 Package는 기존의 TSOP 대비 7배, 가격은 약 10배 정도 높았고, 공급은 삼성전자와 일본의 NEC가 전부였다.

지금은 공정이 발전하여 16Gb 범용 디램이 웨이퍼당 2,000개 이상 양산되기에 500개 수준에 그치는 HBM은 메모리보다는 비메모리에 가깝다는 의견도 있지만, 2000년 당시 128Mb와 256Mb는 당시 공정 (15~19 마이크론)에서 200~400개 수준을 양산할 뿐이다.

Rambus 디램이 양산되며 기존 범용 디램의 Capa를 흡수한다는 것과 Rambus 디램의 System validation이 쉽지 않다는 기록도 있다. 당시 Rambus 디램의 범용 디램 대비 Die Penalty (칩사이즈가 커서 웨이퍼당 생산 칩의 수치가 줄어든다는 의미)는 20%였고, 수율은 10% 수준 떨어져 약 30%의 공급 하락 효과가 있고, Rambus DRAM이 세계 디램 시장의 20%를 차지하면 전체적으로 약 6%의 공급 감소 효과가 발생한다는 분석도 있었다. 현재 HBM의 공급 감소 효과를 설명하는 논리와 아주 비슷하다.

#### HBM과 Rambus

	HBM	Rambus
기존 제품 대비 후공정 비용	5배	7배
기존 제품 대비 초기 가격	6~7배	10배
웨이퍼당 생산량 (Net Die)	500	200~400

자료: 삼성증권 정리

**2000년 IT 버블 붕괴의 원인: 소비자의 외면**

필라델피아 반도체 지수는 2000년 7월 17일 1,266에서 8월 4일 921로 27% 급락하였고, 이는 7월 한 애널리스트의 다운그레이드로 시작되었다. 문제의 리포트 내용은 간단하다. “공급은 늘고 있고, 수요는 선취매와 오버부킹이 문제라서 2~3개 분기 뒤에는 공급 과잉이 발생할 수 있다”는 경고였다.

마이크론의 경우 1998~1999년 2개년 적자를 본 뒤 1999년 4분기부터 흑자 전환 하였고, 때문에 당시의 경고는 너무 이르다는 지적이 많았다. 당시 9월 메모리 업체들도 모든 것이 순조롭다고 언급하였지만, 64Mb 디램 가격은 3분기 8불이 4분기 6.5불로, 다음해 1분기 6불 이후 80% 가량 하락하였다.

**소비자가 외면하는 비용 대비 낮은 효율.** 이 같은 사이클의 붕괴는 대부분 잘못된 전망에 기인한다. 전망이 틀렸다는 것은 또한 소비자가 비용 대비 효율이 높지 않다고 판단한 것에 따른 것이다. 90년대 말 인터넷 열풍으로 초고속 인터넷망이 깔리고 IT 업계는 “보다 많은 사람이 보다 저렴한 인터넷 기기를 사용하게 하자”라는 목표 아래 PC의 원가를 혁신하는 프로젝트에 열중하였다. 일례로 삼보컴퓨터의 20만엔대 초저가 PC가 일본에서 돌풍이었고, 시장 점유율 1위 NEC에 이어 2위까지 성장하고 있었다. 또한 실리콘 밸리에서는 되도록 많은 하드웨어를 하나의 반도체로 구현하여 혁신적으로 원가를 줄이고 성능을 개선하자는 흐름이 있었다.

**2000년 10월 기사 (1): 반도체 고정거래값의 급격한 하락****64MD램 고정거래값 한달새 1달러 하락**

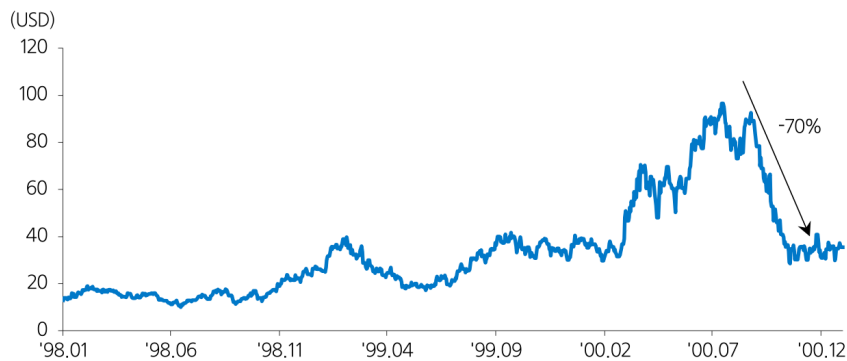
(2000년 10월)

D램 가격이 현물가에 이어 고정거래값도  
계속해서 하락하고 있다.

...

D램 공급 물량의 지나친 증가와 PC 시장 침체가  
반도체 가격의 주요 원인으로 분석된다.

자료: 언론 보도

**마이크론 주가 추이**

자료: Bloomberg

**공격적인 개발 목표의 백지화.** 인텔도 이 같은 흐름에 예외는 아니었다. 인텔은 “팀나”라는 초저가 PC 생산계획을 2000년 10월 백지화한다. \$500 미만의 저가 PC 시장은 당초 예상과는 달리 소비자의 기호가 전자메일과 같은 단순 정보전달이나 검색에서 멀티미디어로 이동하면 외면 받았기 때문이다.

“더 이상 소비자는 저가라고 사는 것이 아닌 제대로 된 PC를 원하기 시작”하였고, 제대로 된 기능을 제공하지 못하는 많은 실리콘 밸리의 벤처들은 금리 인상, 유가 상승 등 경제 환경의 변화에 따라 도산하였다.

특히 인공지능에서 고객의 집중도가 높은 것이 우려될 수 있다. 과거 PC환경에서 Top 3 고객의 주문은 전체 주문의 약 30~40%를 차지하였고, 이후 2016년 클라우드 열풍에서 아마존 등 Top 3 고객의 주문은 전체의 60~70%를 차지하였다. 현재 AI에서 엔비디아가 차지하는 비중은 훨씬 높다. 따라서 엔비디아의 주문이 변한다면 반도체 수급은 갑자기 변할 수 있다.

2000년 10월 기사 (2): 인텔의 초저가칩 생산 계획 취소

### 인텔 초저가 PC칩 ‘팀나’ 생산 계획 무효화

(2000년 10월)

인텔이 수년간 개발해오던 초저가 PC칩 ‘팀나’ 생산 계획을 무효화했다.

...

IDC는 500달러 이하 PC 시장은 이제 사양세이며,  
전자우편만 있으면 만족하던 소비자들이 지금은 고사양을 필요로하는  
멀티미디어 제품만 찾고 있다고 분석했다.

자료: 언론 보도

**디바이스는 혁신의 캐즘에 빠질 수도**

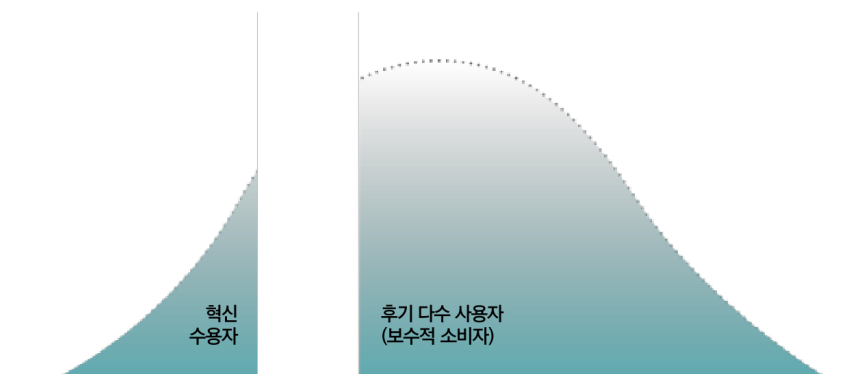
인터넷 버블 이후 스마트폰 시대는 6~7년 이후에나 시작. IT 버블 이후 의미 있는 소비를 이끌어낸 디바이스는 아마도 2002~2003년 와이파이 기술을 달고 나온 노트북일 것으로 생각한다. 소비자들은 이후 커피숍에서도 인터넷을 즐기게 되었다. 하지만 정작 아이폰이라는 스마트폰 시대로 접어들기까지는 그 이후 4~5년의 시간이 더 필요하게 된다.

현재도 인공지능을 위한 수요는 증가하지만 전통적인 개인용 디바이스의 수요가 감소하며 전체 메모리 수요는 정체될 수 있다는 우려가 있다. 디바이스가 아니라 서버가 복잡한 작업을 대신해 줄 수 있기 때문에, 굳이 고성능 디바이스를 자주 구입해야 할 이유가 줄었고, 이는 디바이스의 컴퓨팅 파워 향상에 필수인 고성능·고용량 디램 수요도 함께 줄 수 있다는 의미다.

물론 이런 이유로 메모리 수요가 정체될 것이라는 주장은 20여 년 전부터 반복돼 온 것이긴 하다. 2000년 전후로, 서버와 연결해 작업을 하기 때문에 고성능이 필요치 않고 그래서 가격도 저렴한 '넷북' 열풍이 불면서, '개인용 IT 디바이스의 메모리 성장은 끝났다.'는 얘기가 나오기도 했다. 메모리 추가 성장에 대한 비판론은 이후로도 몇차례 계속 등장했지만 실현되지는 않았다. 실현되지 않은 이유는 서버와 넷북 간의 통신 속도가 충분하지 못했던 탓에 넷북의 소비자 만족도가 예상보다 떨어졌기 때문이다.

그러나 20년간 계속되어 온 이런 주장이 이번에는 현실화될 가능성이 있다. 현재의 5G 무선통신은 20년 전과는 비교도 할 수 없을 만큼의 스피드를 자랑한다. 소프트웨어 즉 AI의 발전과 이를 뒷받침하는 컴퓨터 하드웨어 구조의 혁신이 합쳐지면서, 서버 성능도 크게 높아졌다. 즉 과거 넷북 보급의 걸림돌이었던 통신과 서버 성능의 한계가 이제 문제될 것이 없다는 것이다.

**아직 개발 중인 기술들.** AR·VR 기기는 머리에 쓰는 헤드셋·안경 형태이기 때문에 남들에게 보여지는, 즉 패션 측면에서 거부감이 존재하는 것이 단점이다. 또한 아직까지는 실내 사용 위주다. 강렬한 태양광을 이기고 실외에서 편하게 사용하려면 지금보다 더 밝은 초소형 디스플레이가 개발돼야 한다. 성능은 더 높으면서 크기는 작고 전력 소모도 적은 프로세서가 필요하며, 에너지밀도가 더 높은 초소형·경량 배터리도 개발돼야 한다. AR·VR 기기가 메모리 수요처로 각광받을 수 있지만, 수요 폭발까지는 시간이 꽤 걸릴 수 있다는 뜻이다.

**혁신의 골짜기**

자료: Aly Juma, 삼성증권 재가공

**장기 Risk: AI를 기점으로 중국발 공급 과잉 시작****현재 미중 갈등을 둘러싼 두 가지 착시 현상과 영향**

과거 PC 혁신이 미국 중심이었다면 현재 AI 혁신은 모두의 혁신이므로 더욱 경쟁이라는 관점이 중요하다. 2024년~2025년을 전망하는 입장에서 시장은 다시 두 가지 착시 현상을 경험하고 있다고 판단하고, 단기간에는 이러한 현상이 수요 전망에는 긍정적으로 작용할 수 있다고 생각한다.

하지만 장기적으로 보면 중국은 인공지능의 중요한 하나의 축으로서 첨단 제품을 제조하는 기능을 확대할 것이다. 이는 기존 첨단 제품을 소비하는 국가에서 수출하는 국가로의 변화를 의미하고, 기존에 첨단 제품을 생산하는 국가에게는 위협이 될 수 있다고 판단한다.

두 가지 착시 중 하나는 미국이 중국을 규제하여 중국의 인공지능 역량이 중국에 머물러 있을 것이라는 것이고 (실제 규제는 어렵다는 의미), 다른 하나는 중국의 첨단 제품 제조 역량이 계속 정체하여 있을 것이라는 것이다 (계속 제조 역량이 성장하고 있다는 의미).

**미국산 반도체에 대한 착시 현상: 미국에서 제조하는 것은 효과적이다.** 미국의 장비와 기술이 완전히 봉쇄된다면 중국의 반도체 자립이 실패하는 것은 자명하다. 하지만 아직도 중국의 제품에는 미국이 규제하는 해외 부품이 사용되고 있고, 미국에서 생산되거나 미국의 기술을 사용한 일본 장비의 최대 수요처는 중국이다. 대표적으로 이번 분기 ASML과 도쿄 일렉트론 매출의 절반은 중국이다. 미국 어플라이드 머티리얼즈나 램 리서치도 마찬가지다. 이러한 사실은 미국의 Governance가 효과적으로 작동할지에 대한 불확실성을 의미한다.

또한 미국이 자국에서 반도체를 제조하는 것도 과연 반도체 제조에 필요한 생태계를 미국이 제대로 구축할 수 있을지도 의문이다. 너무 비싼 비용을 감안하면 90년대 미국에 진출했던 일본기업이 미국 진출이 가장 잘못된 의사 결정이었다고 후회한 내용을 상기할 필요가 있다. 그만큼 미국 고객과 접점에서 비메모리를 중심으로 미국에서 반도체를 제조하는 것이 필수적이겠지만, 비용이 높은 미국에서 반도체를 제조하는 것이 호의적인 조건은 아니라는 판단이다.

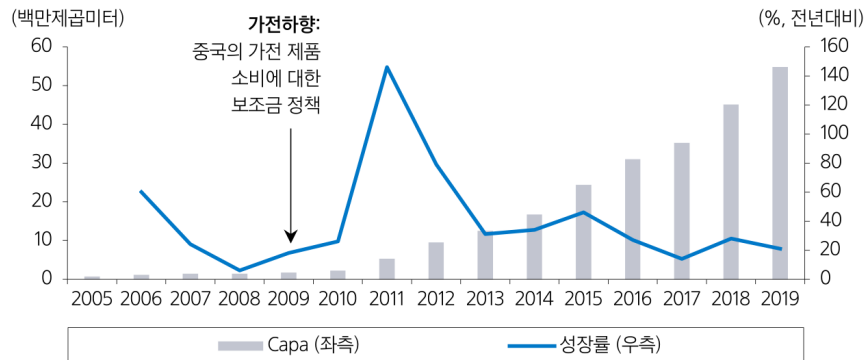
**중국 경제에 대한 착시 현상: 2025년부터 중국발 첨단 제품의 공급 과잉 시작.** 중국 경제에 대한 시각은 너무 부동산 위주로 편향되어 있다고 판단한다. 중국 경제에 대한 우려가 우세하지만, 오히려 중국의 우선순위는 부동산 문제를 해결하는 것 보다, 노동력을 기반으로 한 산업을 첨단산업으로 전환하는 것에 있다고 생각한다. 중국은 반도체 자립을 위해 많은 노력을 기울여왔고, 그 대상은 AI다. 그렇다면 중국을 보는 시각은 이제 첨단 제품의 추과 수요에서 초과 공급으로 바뀔 시점이 되었다는 것이고, 내년부터는 중국발 첨단 산업의 공급 과잉이 시작된다고 할 수 있다.

**향후 수년간 중국의 첨단산업 제조 육성은 공급보다는 수요를 촉진.** 이러한 중국의 첨단 산업 제조를 육성하는 과정은, 우선은 중국 수요 촉진을 중심으로 모두에게 기회로 작용할 것으로 생각한다.

중국이 생산을 시작하려면 우선은 중국 내 소비가 보장되어야 하고, 여기에는 많은 생산과 소비에 대한 보조금이 필요하다. 부동산 문제와 함께 소비가 부진한 것은 사실이기 때문이다. 중국이 자금을 조달하기 위한 국채발행이 가져올 외환에 대한 영향은 제쳐두더라도, 세계 시장의 1/3을 차지하는 IT 제품 소비 시장이 살아난다는 것은 모두에게 호재이다. 현재 AI 서버를 중심으로 한 인프라 수요는 강하지만 모바일 등 전통적인 IT 수요가 나뉘고, AI를 기반으로 한 새로운 기기의 수요는 아직 멀었다는 인식이 내년에 틀릴 수 있다고 생각된다.

아래의 차트는 중국이 Flat panel display 제조를 육성하기 위해 2009년 실시한 가전하향 (가전제품 소비에 판매 장려 보조금을 주는 정책) 이후 급속도로 성장한 중국 패널 업체 BOE의 Capacity이다. 장기적으로 패널 시장에 공급 과잉을 몰고 왔지만 초기 시장 성장에 큰 동력이 되었다.

## 가전하향과 BOE Capacity



자료: IHS

## 잘못된 전략적 판단의 결과

포터는 산업 환경에 영향을 미치는 다섯 가지 요인으로 신규 진입의 위험, 공급자의 협상력, 구매자의 협상력, 대체재, 산업 내 경쟁을 선정했다. 수요에 대한 잘못된 전망과 이를 기반으로 한 어긋난 공급과 투자가 사이클을 만든다면, 경쟁사와 기술에 대한 잘못된 판단은 기업의 생존을 좌우하기도 한다.

**일본과 대만의 경쟁 상태에 대한 오판.** 우선 가장 큰 오판은 아마도 일본과 대만이 한국 반도체 기업의 경쟁력을 평가 절하했던 것이다. 90년대 말을 지나고 일본은 한국이 따라오고 있지만 품질과 기업의 지속 가능성이 떨어져 일본이 버티면 이긴다고 하였다. 하지만 쌓여가는 적자와 고정비 지출에 대한 부담을 극복하지 못하고 몰락하게 된다. 2007년 금융위기로 기업실적이 악화되자 대만은 미국의 기술과 대만의 원가를 합치면 한국을 이길 수 있다고 자신하였다. 하지만 PC 위주의 공정이 주력이었던 대만은 모바일 환경에서 경쟁적인 원가를 확보하지 못했고, 역시 몰락하게 된다.

**중국의 기술과 자국 시장에 대한 오판.** 기술과 시장에 대한 오판은 더욱 치명적인 결과를 초래한다. 인터넷이 보급되고 모두가 인터넷의 수요를 장악하는 새로운 엔진을 공급하겠다는 포부를 밝힐 때의 일이다. 필립스는 새로운 엔진이 TV가 될 것이라고 판단했지만, 시장은 보다 작은 스마트폰에서 형성되었기에 LG와의 협력은 예상과 달리 성공적이지 못하였다. 스마트폰 시장 진입에 실패한 노키아의 몰락도 같은 배경이다.

중국이 오래 전부터 반도체 시장에 진출하려 하였지만 아직도 의미 있는 성과를 내지 못하는 것은 미국의 규제가 결정적이지만, 중국이 당초 가졌던 전략이 잘못된 것일 수도 있다. 중국의 전략은 여러 지방성이 제각기 반도체를 추진하여 그중 성공하는 곳에 모든 자원을 몰아주는 방식이었다. 기존에 중국이 성공하였던 제조업과 달리 반도체는 투입되는 자본도 비교할 수 없을 만큼 크지만, 그만큼 의사결정의 신속성과 투명성이 필요한 곳이다. 중국은 뒤늦게 반도체와 관련된 부패를 줄이고 중앙정부 위주의 의사결정 구조를 가져가지만, 그만큼 귀중한 시간이 허비되는 동안 경쟁국의 견제는 높아졌다고 판단한다.

**중국 반도체: It is crazy**

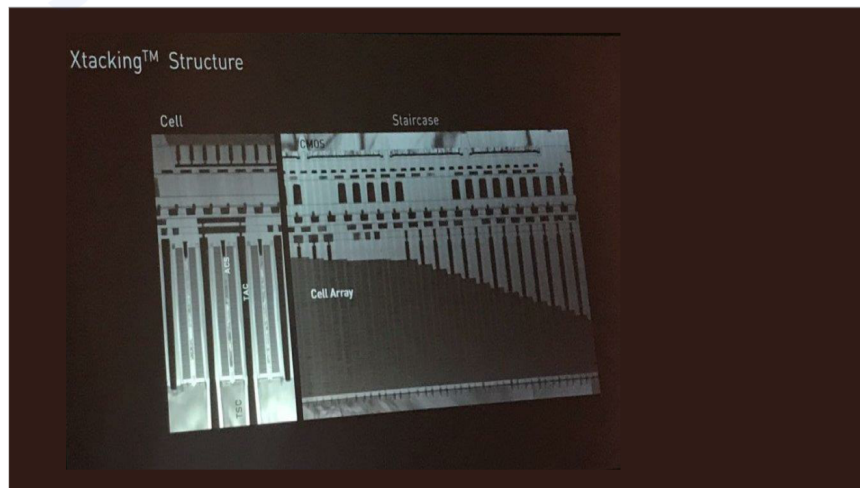
전문가의 경계심은 낮아지고, 중국의 목표는 더욱 커지고 있다. 인공지능의 혁신은 반도체의 혁신이 필수적이기에, 미국이 중국의 반도체를 규제한다면 중국의 인공지능은 경쟁력을 가지기 어려울 것이라는 의견이 가능하다. 고부가 가치의 GPU와 EUV 등 신규 장비를 규제한다면 이러한 의견은 더욱 지지를 받는다. 실제로 지난 수년간 중국 반도체에 대한 전문가들의 경계심은 낮아지고 있다고 판단된다.

하지만 중국의 반도체 증설 계획은 실제로는 더욱 커지고 있는 것이 사실이다. 예를 들어 디램을 제조하는 CXMT의 경우 기존에는 향후 5년안에 성공적으로 성장을 해봐야 생산 규모가 현재의 마이크론 수준의 월 30만 장 규모에 그칠 것이고, 이것도 아주 낙관적이라는 의견이 지배적이다. 하지만 중국의 목표는 이보다 2배가 큰 월 60만 장 규모로 더욱 커지고 있다.

이 같은 증설 계획은 합리적이지 않거나 Crazy하다고 할 수 있지만, 과거 한국이 가졌던 “합리적이지는 않지만 투자를 늘리고 기술을 개선하면 점유율을 확보할 수 있다는 단순한 믿음”과 동일하다. 그리고 이러한 믿음은 중국 시장의 점유율을 높게 가정하는 것에 기반한다. “왜 중국 반도체 업체가 중국시장을 100% 먹는다는 생각을 하지 않는가”라는 질문이 우습게 들릴 수 있지만, 정부의 보조금을 기반으로 당연시되는 고객의 Validation 등 절차를 무시하고 진행되는 증설은 거침이 없다.

이가 없으면 잇몸으로. EUV가 없으면 당장 확보된 기존 장비에는 한계가 있고, 그나마 기존 장비도 유지보수가 어려우면 사용이 어렵다는 어려움도 있다. 하지만 이가 없으면 잇몸으로 때운다는 말처럼 중국은 차선을 찾아 도전하고 있다. 예를 들어, 낸드 플래시에서 미세화의 장벽에 도달하여 시도한 Xtacking (낸드의 미세화가 어려워 두 개의 웨이퍼에 메모리와 로직다이를 별도로 제조하여 서로 본딩하는 방법)이 있다. 모두가 얼마나 미세화가 되지 않으면 웨이퍼 두개를 붙여 사용하느냐고 하였고, 당연히 상승하는 원가부담으로 얼마 가지 못할 것이라고 하였다. 하지만 Xtacking 기술은 현재 모두가 사용하는 기술이 되어가고 있다.

Xtacking 기술은 낸드에만 적용되는 것이 아니라 디램에도 적용될 것으로 예상된다. EUV가 없으면 중국은 미세화가 어려워 3D 디램을 추진하고, 이때 셀을 3D로 만든 다음 셀과 페리를 따로 제조하여 둘을 접합하는 식이다.

**3D Xtacking**

자료: YMTC



**CXMT 2025 상장과 중국 증시를 통한 자금 조달 계획**

**적자라도 상장.** 중국 당국은 CXMT가 생산 원가(Cash Cost)만 감당할 수 있다면 상장을 시켜준다는 방침이다. 작년 기준 CXMT의 손익은 분기에 10억 달러 수준의 적자로 추정된다. 60억Gb 생산·판매에 Gb당 가격을 0.2달러로 가정하면, 매출은 12억 달러다. 분기 적자 10억 달러, 연간 적자 40억 달러로 가정하면 원가는 52억 달러다. 그러면 개당(Gb 기준) 원가는 0.82달러가 되고, 변동비를 총비용의 3분의1(0.3달러)로 계산하면 고정비는 개당 0.5달러가 된다. 이렇게 되면 변동비 18억 달러를 감당 못 하게 되지만, 변동비를 개당 0.2달러로 산정하면 이익이 가능하다. 실제 2024년에는 범용 디램 가격이 오르며 Gross margin으로는 흑자 전환이 되었다는 판단이다.

**장비와 부품이 없다면? 백장갑과 조끼라는 구매와 주문 대행사들.** 미국이 더욱 중국에 대한 반도체 장비 규제를 강화할 수 있기 때문에 중국은 2025~2026년까지 사용할 장비를 이미 비축하였다고 판단된다. 그 이후는 국산화가 필요하지만, 성공한다면 앞서 설명한 점유율 확대가 가능하다.

중국에는 해외 장비나 부품을 대신하여 구매하는 “백장갑”으로 불리는 중간 유통 회사들이 있고, 파운드리에 제조를 위탁하는 것도 “조끼”를 입었다는 의미로 중간 유통 회사들을 사용하고 있어 실제 장비와 부품이 중국 누구에게 가는지 추적이 어렵다. 중국의 인공지능 시스템을 열어보면 외국산 반도체가 어렵지 않게 보이는 것도 때문이고, 물론 최대 수요처는 화웨이이다. 중국은 이들 인공지능 반도체도 향후 2-3년 안으로 국산화한다는 방침으로 개발이 진행되고 있다.

**그러면 중국 반도체 주식을 사야 할까? 보다 유연한 인력 확보에 대한 확인 필요하다**

PwC의 의견대로 인공지능 시장의 성장으로 가장 수혜를 보는 지역이 중국이고 중국의 인공지능 반도체 육성이 거대한 시장을 중심으로 거침없이 진행된다면, 우리는 중국의 반도체 주식을 사야 할까?

아마도 중국 반도체 업체들은 높은 이익을 내기는 어렵겠지만, 높은 매출 성장이 가능할 수 있다. 설비 투자에 대한 보조금과 판매 보조금을 바탕으로 낮은 비용을 가져간다면 이익 성장도 가능하다. 또한 신규 상장이 이루어진다면 국가의 지원을 바탕으로 인기몰이도 가능하다.

그러나 아직 해결해야 할 점이 있다. 다시 강조하는 것은 인공지능 반도체는 높은 성능이 필수적이라는 점이다. 하지만, 아직 중국이 보유하고 있는 반도체 공정은 전하의 이동 속도가 기존 업체들에 비해 느리다. 그것을 따라잡으려면 다시 반도체 노하우의 문제로 돌아간다. HKMG (High K Metal Gate) 등 물질을 다루는 소재와 구조의 노하우가 어려운데, 이를 해결하는 것은 결국 사람이다.

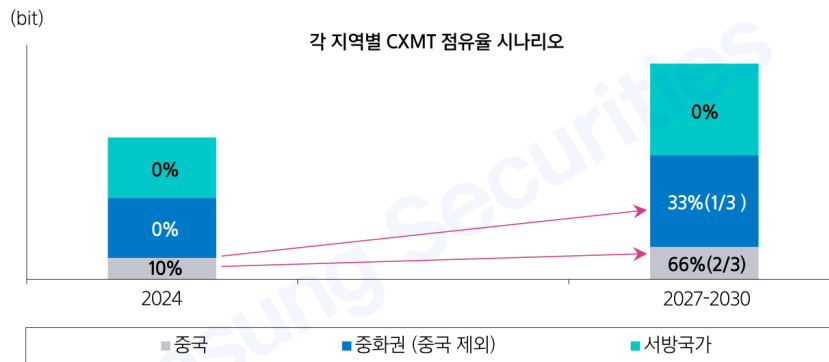
문제는 중국 내 분위기가 외국 인력을 배척하는 분위기로 바뀌었다는 점이다. 2016년부터 중국이 반도체를 육성하는 초기에는 외국 기술진에 많은 보상과 권한을 주었다면, 이제 중국 반도체 기업들은 보다 중국의 내부 인력을 위주로 구성되고 있다. 워낙 미국의 규제가 민감하기도 하고, 부패 등 문제가 많았던 지방성 중심의 관리체계를 중앙정부와 화웨이 등 대표적인 기업을 중심으로 전환하며 보다 통제가 강화된 분위기다.

오히려 반도체 노하우에 대한 습득과 축적은 외국인력의 노하우와 복제를 통한 반도체 육성 초기보다 느려질 수 있다는 판단이다. 물론, 인력이 통제될 때 지나친 목표에도 과장된 자신감을 표현하는 가능성도 배제할 수 없고, 이는 계획대비 양산이 지연될 수 있다. 때문에 중국 반도체에 대한 투자는 이러한 인력 확보에 대한 유연성을 먼저 확인할 필요가 있다는 판단이다.

**중국의 반도체는 모바일과 가전 시장부터 진입**

중국 반도체에 대한 투자는 이러한 인력 확보에 대한 유연성을 먼저 확인할 필요가 있다는 판단이지만, 그럼에도 불구하고 우리가 장기적으로 중국 반도체에 대한 경계를 높혀야 하는 것은 중국의 거대 시장과 이를 배경으로 산업 고도화를 위한 정책적인 지원, 그리고 여기에 매진하는 많은 기술 인력 때문이다.

**중국 시장의 2/3, 주변 국가의 1/3.** 기술 발전 속도가 느리지만 중국은 계속 도전하면 결국에는 중국시장을 점유하고 미국을 제외한 주변의 국가들에게 수출할 수 있다는 생각이다. 중국 시장은 전 세계 시장의 약 1/3이다. CXMT의 중국 내 점유율이 현재 약 10~20% 수준이라고 보면, 2030년 이전에 2/3에 도달하겠다는 생각이고, 미국 이외의 해외 점유율도 현재는 없지만 같은 기간에 1/3 이상을 목표할 수 있다. 가장 손쉽게 2/3 수준의 점유율을 확보할 수 있는 시장은 가전과 모바일 시장이다. 인공지능을 포함하는 서버 시장은 1/3 수준의 점유율을 확보하는 것이 우선이다.

**중국 디램 Capa 시나리오**

자료: 삼성증권

**화웨이 Pura70의 사례.** 화웨이의 최신 스마트폰 Pura70에는 12GB와 16GB 버전이 있고 12GB가 메인이다. 12GB 버전에는 중국의 CXMT가 공급하며 12Gb를 8개 사용하여 만든 것이다. 미세화가 뒤진 공정 기술로는 속도와 소비 전력에서 현재 삼성이나 애플의 스마트폰보다 성능이 떨어진다. LPDDR5는 4266MHz의 LPDDR4에서 50% 더 빨라졌다. 이제 LPDDR5는 8533이나 9600MHz로 더 빨라지고 있지만, 중국의 디램은 아직 6400MHz에 머물러 있다.

속도가 떨어지는 것은 중국이 아직 HKMG 기술이 확보되지 않았고, 화웨이의 AP도 뒤떨어진 공정에서 양산되어 높은 속도를 지원하지 못하기 때문이다. 중국이 내년까지 HKMG 등 공정을 확보하면 속도를 높이고 소비 전력을 줄일 수 있지만 노후를 얻기 어려워 쉽지 않다. 인텔과 같은 업체들의 validation 등을 무시하고 정부의 보조금으로 진행되는 사업은, 불량 등 소비자에 대한 사후지원도 어렵다.

객관적으로 보자면 화웨이의 인공지능 스마트폰은 외국의 첨단 반도체를 구매하여 사용하지 않는다면 성능이 떨어지거나 과대광고에 가깝다고 할 수 있다. 하지만 중국의 대표기업인 화웨이의 최신 AI 스마트폰에 중국산 메모리가 탑재되기 시작하였다는 점은 시사하는 바가 크다.

**중국은 2023년에도 감산은 커녕 증설**

역대급 재고 탓에 한국의 2023년 디램 생산은 전년 대비 20% 이상 줄어들었다. 재고를 소진하려면 신규 생산을 줄이는 방법밖에 없기 때문이다. 하지만 중국 디램업체 CXMT는 감산은 커녕 20% 가량 증산을 하였다. 생산 규모는 2022년 50억Gb, 2023년 60억Gb로 600~900억 Gb(삼성 900억Gb, SK하이닉스 600억Gb) 수준의 한국과 큰 차이를 보이지만, 작년과 같은 사상최악의 불황에도 오히려 생산을 늘렸다는 것이 충격적이다. 중국이 불황에도 메모리 증산이 가능한 이유는 국내 수요가 워낙 크기 때문이다. 중국 메모리 회사가 더 많이 만들어도, 중국 내에서 소화 가능하다는 얘기이다.

**중국의 곡선 구간 이론과 업계의 대응**

물론 중국이 해결해야 할 과제도 많다. 중국은 미국의 대중 수출규제 탓에 EUV 등 첨단 제조장치를 쓸 수 없다. 따라서 상대적으로 성능이 떨어지는 중국 장비를 더욱 많이 사용해야 하고 독자개발에 더 나서야 한다는 문제가 있다. 중국의 의사 결정 구조도 개선해야 한다. 반도체는 기술이 어렵고 자본투자가 크기에 중국이 성공한 다른 산업에 비해 훨씬 더 투명하고 효율적이어야 하기 때문이다.

**중국의 곡선 구간 이론.** 그러나 기회도 병존한다. 현재의 디램 개발은 기술이 급속도로 발전하던 시기를 지나 일종의 기술 정체기를 겪고 있다. 이런 시기엔 집중투자가 가능하고 시장의 힘도 활용할 수 있는 중국이 추격하기 쉬워질 수 있다. 이것을 중국에서는 '곡선 구간(커브) 이론'이라 부르기도 한다. 메모리 반도체 산업의 특성을 자동차 경주에 빗댄 것이다.

직선주로는선 웬만큼 가속하더라도 앞서가는 차를 따라잡기 쉽지 않다. 상대를 잡을 수 있는 기회는 앞서 가는 차도 속도를 줄일 수밖에 없는 곡선 구간이다. 이 곡선 구간이란, 디램의 업황이 안 좋고 기존 기술이 정체돼 있는 시기를 뜻하는 것이다. 중국의 생각은 지금이 바로 곡선 구간이며, 이 구간에 있을 때 전력을 다해 추월해야 한다'는 것이다. 중국이 곡선 구간에서 앞차를 잡으려면, 이때 치고 나갈 수 있는 차별화된 능력이 필요하다. 3D 메모리 기술이 그 사례이다.

물론 3D 기술은 한국 업체들도 파악하고 있고, 굳이 뒤따라오는 차량이 구사하려 하는 기술을 똑같이 사용해 추격을 막아야 한다고 생각하지 않을 수 있다. 아직은 앞선 자의 여유다. 그러나 모두가 곡선 구간에 진입한다면, 앞선 자가 여유를 부리기 어려워질 수 있다. 중국은 이 점을 이용하려는 것이다. 앞선 경쟁자가 곡선 구간에서 주춤하고 있는 틈을 노리겠다는 것이다.

**중국산 7나노 로직반도체, 그다음은 3D 디램**

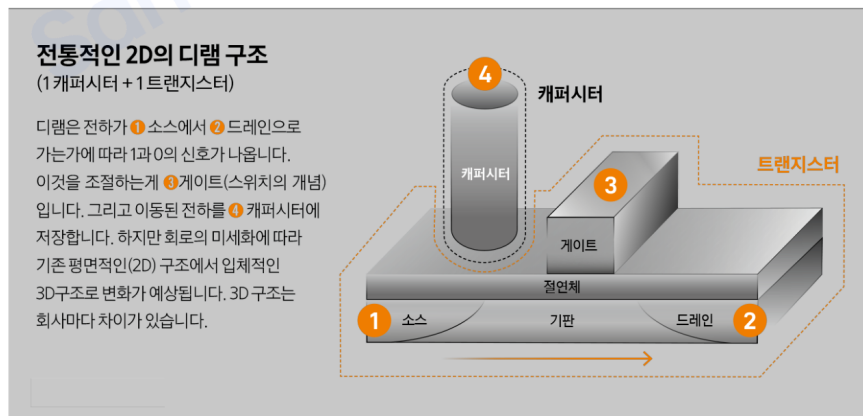
**화웨이와 SMIC의 7나노 파운드리, 다음은 디램.** 중국은 7나노 반도체 독자생산 선언을 시작으로 미국의 규제를 벗어날 수 있다는 사례를 추가적으로 보여줄 것으로 예상되고, 그 대상은 디램이 될 것으로 판단한다.

화웨이가 중국 파운드리인 SMIC의 7나노 공정을 활용하여 화웨이의 신규 스마트폰 AP를 조달한 것이 화제였고, 이후 화웨이는 단숨에 애플을 제치고 중국 스마트폰 시장 1위를 되찾았다.

미국은 로직 반도체 7나노, 디램 18나노, 낸드 128단을 포함하여 이보다 미세한 공정에 활용되는 미국 장비가 중국으로 수출되는 것을 규제하고 있다. 7나노 로직 반도체를 만들려면, 네덜란드의 ASML이 100% 독점 제공하고 있는 EUV 노광장비를 써야 하는데, 미국이 EUV 장비의 중국 수출을 막고 있다. 그런데 중국은 EUV를 사용하지 못하는 대신, EUV보다 구세대인 DUV(심자외선) 노광장비를 사용하여 양산 기술을 확보한 것이다. 구형장비를 사용한 탓에 성능과 원가경쟁력 등이 모자라고 생산규모도 제한적일 수 있지만, 중국이 미국의 반도체 규제를 피하여 독자 생존할 수 있음을 보여줬다는 것이 의미가 크다.

**3D의 파괴적 혁신.** 7나노 독자생산은 이제 시작일 수 있다. 중국은 미국의 규제를 벗어날 수 있다는 사례를 추가적으로 보여줄 것으로 예상되고, 추가 사례는 디램이 될 것으로 판단한다. 기존의 2D 디램이 평면에 셀을 쌓는 것과 달리, 3D 디램은 수평·수직 양방향의 입체(3차원)로 셀을 쌓아 저장 용량을 높일 수 있다. 이렇게 되면 회로 선평의 미세화에만 초점을 맞추지 않아도 성능 향상이 가능하고, 디램 회로의 선평(18나노)에 초점을 맞춘 미국의 규제를 피해갈 수 있다.

중국이 미국의 회로 선평 규제에 막혀 구형의 2D 디램만 제조할 수 있다면, 한국이 크게 걱정하지 않아도 된다. 그러나 중국이 3D 디램의 양산에 성공한다면 상황은 완전히 달라질 수 있다. “중국이 한국의 첨단 메모리를 추격하는 것은 어렵다. 따라오더라도 중저가 일부제품에 그칠 것”이라는 전제가 단번에 무너질 수 있다.

**전통적인 2D의 디램 구조 (1 캐패시터 + 1 트랜지스터)**

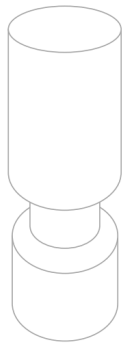
자료: SK하이닉스, 삼성증권 재가공

## 3D 디램 구조



자료: Applied Materials

## CXMT 3D의 디램 구조



## 트랜지스터를 원통형으로 구성

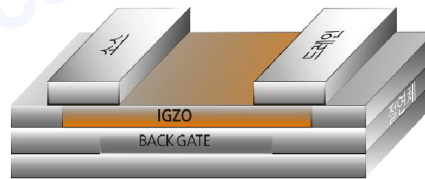
미세화로 인해 트랜지스터에서 누설전류와 간섭현상들이 발생하는것이 디램의 문제입니다. 트랜지스터를 위로 쌓으면 트랜지스터 간격이 넓어서 문제 해결이 가능합니다.

자료: CXMT, 삼성증권 재가공

## 화웨이의 또다른 3D 디램 구조

## 캐패시터 없이 두 개의 트랜지스터로 구성, 이그조(IGZO)를 활용한 점이 특징

트랜지스터를 쌓아 미세회로의 부담을 덜어도 캐패시터의 미세화 부담은 그대로입니다. 그래서 캐패시터를 없애고 트랜지스터 두개로 전환하는 겁니다. 특히 IGZO는 낮은 대기전력이 장점으로 디램에서 새로운 물질로 여겨지고 있습니다.



자료: imec, 삼성증권 재가공

**화웨이의 여러가지 시도가 가능한 배경은 시장과 보상.** 화웨이가 3D 디램을 제조하는 방식은 2개의 트랜지스터를 활용하는 것도 있다. 기존 디램은 캐퍼시터(capacitor: 전류 저장소) 1개와 트랜지스터(꺾다 켜는 스위치) 1개가 한 세트인데, 캐퍼시터를 빼고 트랜지스터를 추가한 것이 화웨이의 3D 디램 제조 방식이다. 업계에서는 단기간에 실현하기 어렵다는 게 일반적이다.

화웨이가 이러한 새로운 3D 디램을 개발할 수 있는 이유는 일본의 원천 기술을 흡수했기 때문이다. 화웨이 방식을 쓰려면 이그조(IGZO, Indium·Gallium·Zinc Oxide, 인듐·갈륨·아연·산소) 디스플레이 기술이 필수인데, 이 기술은 일본이 갖고 있다. 화웨이는 일본의 기술뿐 아니라 인력도 함께 흡수하였다. 화웨이는 일본 요코하마 연구소를 세우고 관련 기술을 개발했던 일본의 마지막 반도체 회사 엘피다(2012년 파산, 미국 마이크론에 인수) 출신 인력과 함께 3D 디램을 개발하였다.

화웨이 같은 중국 기업이 일본 인력을 활용할 수 있는 이유는 충분히 보상하기 때문이다. 일본 국적의 메모리 회사가 사라졌다고 해서, 관련 기술과 기술자까지 사라지는 것은 아니다. 제대로 보상하고 시장 가능성을 제시한다면, 일본 기술자가 중국과 협업하지 않을 이유가 없다. 물론 화웨이가 이러한 기술 개발 계획은 현실적 대안에서 멀어지고 있다. 산업계는 보다 간단한 CXMT의 수직적인 3D 구조를 대안으로 추진하고 있다.

**CXMT의 디램.** 디램 전문업체 CXMT도 3D 디램 개발을 진행하고 있다. 방식은 일반적인 업계의 방식으로 트랜지스터를 삼성 파운드리 3나노의 GAA(Gate All Around) 공정처럼 원통으로 쌓고, 캐퍼시터는 원통형 트랜지스터 위에 올리는 방식이다. 2026년 개발이 목표인데, 어려운 목표이지만 대안이 없는 상황에서 적극적으로 추진되고 있다.

CXMT의 현재 주력 공정은 이 업체 기준의 3세대(Gen3)로 삼성의 19나노(1x)급이다. 삼성전자는 1x 이후 1y, 1z를 거쳐 1a가 주력이기 때문에, 한국보다 3세대 뒤쳐진 것이다. 선도 기업대비 5년 이상 뒤쳐진 기술이라 볼 수도 있다. CXMT는 자사의 3세대를 17나노라 부르지만 실제로는 한국의 19나노급이다.

CXMT의 주력 상품은 8Gb LPDDR4로 중저가 스마트폰에 장착되고 올해는 12Gb LPDDR5로 전환되고 있다. 8Gb LPDDR4는 이미 옛날 제품이라 한국에 부담이 되지 않는다는 반론도 가능하다. 그러나 기술 수준이 낮은 반도체에 대한 중국 내수요가 여전히 높고, 그 수요를 정부와 기업이 합작해 늘릴 수 있다는 게 중요하다.

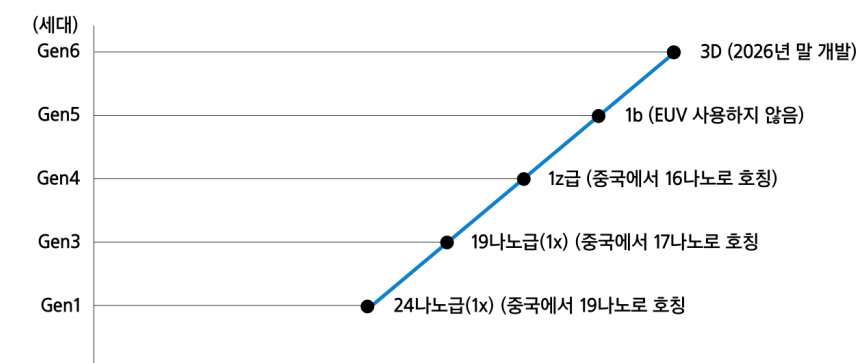
중국 스마트폰 업계는 성능이 다소 떨어지더라도 중국산 메모리를 쓰는 방향으로 제품 설계를 바꿔 나가고 있다. 기술적으로 개선된 POP(Package On Package)나 MCP(Multi Chip Package) 방식 대신, 이보다 수준이 낮지만 중국 메모리업체의 방식인 디스크리트(Discrete) 형태의 모바일 디램(LPDDR4)을 사용하는 것이다.

POP은 프로세서 안에 메모리를 내장하는 방식으로, 애플 아이폰에 사용된다. MCP는 2개 이상의 반도체 칩을 쌓아 하나로 만드는 방법이다. 예를 들어 디램과 낸드를 하나의 패키지로 만들어 스마트폰에 부착하는 식이다. 개별 반도체라고도 하는 디스크리트는 제품 내에서 단일한 기능을 하는 부품이다. POP나 MCP처럼 쌓지 않고 디램이나 낸드를 각각 기판(PCB)에 붙이는 것이다. 디스크리트는 POP·MCP에 비해 기판 면적을 많이 잡아먹기 때문에 주로 저가형에 사용된다.

중국은 2027년까지 중국에서 사용하는 모든 메모리를 국산화한다는 방침이다. 2025년 말까지 자체 AI 서버를 구축한다는 계획인데 AI 서버에 들어갈 16GB HBM3도 같은 해 양산한다는 목표이다.

중국의 디램 생산능력이 아직은 작으나 증산 속도가 빨라 곧 위협이 될 수 있다. CXMT의 허페이 1공장은 웨이퍼 기준 월 9만 장, 북경 1공장은 월 9만 장의 규모로 증산이 가능하다. 허페이 2공장은 내년 양산(Gen 4)을 목표로 하고 있고, 한국의 1z급이다. Gen 5는 마이크론 기준 1b(마이크론이 EUV 없이 한 것을 모방한 것으로 판단)로 2026년 말까지는 생산할 가능성이 있다. 3D 디램은 Gen 6로 역시 2026년 말 개발 완료 목표이다.

#### CXMT의 세대별 공정 기술



자료: CXMT, 삼성증권 정리

**미·중 갈등의 종착점은 합의, 합의의 초점은 지식 재산권과 보조금**

미국과 중국이 결국에는 일정 수준의 합의에 이를 것이라 가정하는 것이 타당하다고 당사는 판단한다. 미국과 중국이 대만을 두고 물리적 전쟁을 하거나, 일부 언론 보도와 같이 미국이 최악의 경우 TSMC의 대만 내 시설을 파괴하는 시나리오도 상정할 수 있지만, 이러한 극단적 시나리오는 미국과 중국 양 쪽에 극심한 타격을 줄 수 있기 때문에 서로가 끝까지 피하려 할 것이다.

**갈등은 장기적으로 조정될 가능성.** 기술적으로 봐도 지금의 극단적 대치는 일정 수준에서 조정될 가능성이 있다. 수십년간 반도체 제조 부문에 힘을 신지 않았던 미국이 일거에 제조 경쟁력을 높인다는 것도, 중국이 미국의 장비와 설계툴(tool 개발이나 설계를 지원하는 범용 프로그램) 등을 쓰지 못하도록 하는 수출 규제를 피해가며 첨단 반도체 경쟁력을 단기간에 획득하는 것도 가능성이 떨어지기 때문이다. 또 기술적 문제뿐 아니라, 미·중을 둘러싼 거시경제 환경 악화가 어쩔 수 없이 합의로 이끄는 원인이 될 수도 있다. 한국에 최상의 시나리오는 미국이 중국의 반도체 진입을 막아주는 것이지만, 그런 시나리오만 기대하기보다는 미국과 중국의 합의를 염두에 두고 그 내용과 시기를 예상해 대비할 필요가 있다.

**지재권과 보조금.** 미국과 중국이 합의해야 할 대상은 ①지식 재산권과 ②보조금이라고 생각된다. 미국 입장에서는 지식 재산권이 보장되어야 중국의 반도체 추격 속도를 늦추고 자신들의 시장을 오래 확보할 수 있다고 생각할 것이다. 중국의 반도체 첨단공정 경쟁력이 아직 떨어지고, 이를 보완하기 위해 미국 등의 지식 재산권을 확보해 나가려면 많은 시간과 비용이 들 것이기 때문이다. 또 미국의 지식 재산권 시장에 중국이 편입됨으로써 미국 반도체 기업이 수익을 늘릴 수 있다. 수익이 늘면 연구개발에 더 투자할 수 있고, 그렇게 미국 내 투자가 더 늘어나 고급 반도체 인력이 미국으로 집중되는 환경을 유지할 수 있다.

**미국이 중국을 보는 법: 기업은 시장을, 정부는 규제를 우선**

 애플	 테슬라	<b>이익 지향</b> - 중국 생산을 차츰 줄일 계획이나, 중국 시장은 포기 못해 - 중국산 부품을 적극적으로 구매
 미국 정부	<b>지금미 규제의 적기(golden time)이다</b>	
 인텔	 엔비디아	<b>중독 전략</b> - 인텔 : 대중국 수출 규제하면, 중국이 5년 안에 반도체 국산화 - 기술 격차를 두고 미국 제품에 중독시키고, 최악의 경우에는 중국 사업 종료
 램리서치	 어플라이드 머티리얼즈	<b>혁신 지향</b> - 미국의 우위는 투자와 글로벌 시장을 필요로 하는 기술에 있음 - 기술 우위는 R&D에 있는데, R&D는 글로벌 시장에서 오는 수익이 기반 - 시장 규제로 수익이 줄면 핵심 엔지니어들이 미국을 떠날 것이고, 제국은 무너질 것

자료: 각 사, 삼성증권



**지식 재산권은 상대적으로 쉬운 협상. 보조금은 어려운 협상.** 지식 재산권에 대해서는 협상 여지가 있어 보인다. 반면 보조금 문제는 협상이 더 어려울 수 있다. 중국 같은 후발주자에게 보조금을 끊으라는 것은 추격의 싹을 자르겠다는 말과 다름없기 때문이다. 특히 반도체 사업은 투자금액이 크게 증가하는 추세이기 때문에, 처음부터 자생력을 키우는 것이 쉽지 않다.

중국의 반도체 보조금도 미국에 중요하다. 중국은 미국·대만·한국과의 기술·생산능력 격차를 단기간에 줄이기 위해, 수익성을 우선하지 않는 집중 투자와 대량 증설에 나서고 있다. 이를 가능케 하는 것이 중국 정부가 자국 반도체 산업에 지원하는 보조금이다. 중국의 보조금을 제한한다면 중국의 추격 속도를 늦출 수 있다.

중국의 신규 메모리 반도체 업체 육성에는 매년 연구 개발에만 최소 1조원이 투입된다. 이러한 자금은 매몰 비용(sunk cost·어떤 선택의 반복 여부와 무관하게 회수할 없는 비용)으로 사라지게 된다. 또 고객에게 주문을 받으려면 적어도 20조원의 자금이 보장되어야 한다. 웨이퍼 기준으로 월 7~8만 장은 되어야 안정적 공급을 보장할 수 있는데, 1만 장 규모의 증설에 2조원 이상 비용이 들기 때문이다. 중국의 경제구조가 지방의 성(省) 위주로 되어 있고 성의 자금이 제한적임을 감안하면, 중앙정부의 보조금 없이는 시작 자체가 불가능한 계획인 셈이다.

**판매 보조금.** 중국의 보조금은 설비 투자와 운영을 넘어 판매에도 적용된다. 이전 중국이 자국 내 패널 산업을 육성하기 위해 사용했던 가전하향(家電下鄉)과 같은 방식이다. 가전하향은 가전제품을 농촌으로 내려 보낸다는 뜻으로 가전제품을 구입하는 농촌 지역민에게 일정한 보조금을 지급하는 정책이다. 2013년초까지 4년간 실시되었고, 중국 패널 산업 육성의 중요한 동력이 되었다.

중국의 반도체 보조금 지급은 이런 식이다. 예를 들어 화웨이가 100원짜리 중국산 디램을 구매하면, 국가가 화웨이에게 20원을 지급하는 식이다. 생산업체 입장에서는 20%의 가격 경쟁력을 덤으로 갖게 되는 셈이고, 처음부터 자생력 가격 경쟁력을 갖기 어려운 중국 디램 업체에는 큰 힘이 된다.

## 부록 1: 인공지능 기술 개요

### 인공지능 기술 발전 역사 및 발전 방향

혁신은 합리적인 계획에 따라 만들어진다고보다 용감하게 하는 과정에서 되어 버렸다는 표현이 더욱 어울린다. 현재 각광받는 AI도 사실 새로운 것이 아니다. 90년대 중반에도 냉장고와 대화하는 음성 인식 기술이 있었다. 머신러닝의 새로운 방법이 도입된 것이 2006년이고 이를 더욱 발전시킨 것이 2017년의 트랜스포머 모델 (All you need is attention)이지만, 이를 성공에 이끈 것은 혁신가들의 과감한 투자와 집요한 노력이었다. 과거에는 상상할 수 없었던 방대한 데이터를 연산하고 이를 통해 학습과 추론을 해야 했기 때문이다.

인공지능 기술을 알기 위해서는 인공지능의 시작이 된 기계 학습부터 개념을 살펴보자. 흔히 착각하는 부분이 인공지능과 기계 학습은 다른 개념이라는 것이다. 그러나 인공지능과 기계 학습은 완전히 상이한 개념이 아니라, 인공지능이라는 가장 큰 상위 개념에 기계 학습이 포함되어있다. 좀 더 고도화된 알고리즘을 형성하기 위해 인공지능은 기계 학습에서부터 시작해서 심층 학습, 강화 학습 등 기계 학습이 구체화되는 방식으로 학습 방법이 발전했다.

#### 1) 기계 학습 (Machine Learning)

기계 학습은 현재 인공지능에서 가장 중추적인 역할을 하는 알고리즘으로, 말 그대로 '기계가 스스로 학습'하는 알고리즘이다. 일일이 코드로 명시하지 않아도 기계가 데이터를 학습하여 미래를 예측하고, 동작을 실행하는 기술이다.

기계 학습 이전의 인공지능은 전문가 시스템에 의존했다. 예를 들어, 동물 분류 시스템에서 동물학자가 직접 생물 분류 규칙을 입력하면 이를 토대로 인공지능이 작동했다. 그러나 전문가 시스템은 인간이 알지 못하는 것은 기계도 알 수 없다는 한계가 있다. 또, 같은 종이어도 동물의 모습이 동일하게 나타나는 것도 아니기에 입력된 규칙으로 모든 동물을 분류하기엔 한계가 있다.

기계 학습은 방대한 양의 데이터를 분류하고 학습시키는 과정을 거쳐 기계가 입력된 규칙 외에도 스스로 동물을 분류할 수 있도록 한다. 예를 들어, 전문가 시스템에선 강아지는 흰색, 검은색이 존재한다고 하고, 그 외 특징을 입력하면, 흰색 혹은 검은색 강아지만 강아지라고 판단할 수 있다. 반면, 기계 학습에선 어떤 동물이 색상 외에 강아지의 다른 모든 특징을 만족하고, 털만 갈색이라면 이 동물이 강아지라고 합리적인 판단을 내릴 수 있다.

단, 기계 학습에도 적용 가능한 범위가 제한적이라는 단점이 존재한다. 기계에 학습을 시키기 위해서는 데이터의 프로그래밍이 필요한데 숫자 혹은 단순 명료한 데이터의 경우 프로그래밍이 가능하지만, 그 외 데이터를 프로그래밍하는 것에는 어려움이 있다. 또, 최대한 많은 특징을 인간이 직접 추려 사전에 기계를 학습시켜야 한다는 한계가 있었다. 즉, 기계 학습 기반 동물 분류 시스템을 활용하려면 인간이 강아지의 특징을 모두 추려 수많은 예시를 통해 기계를 학습시켜야 한다.

## 2) 심층 학습 (Deep Learning)

기계 학습의 한계점을 해결하기 위해, 인간의 두뇌를 소프트웨어로 재현하는 방식인 신경네트워크에 관한 연구가 진행되었다. 신경네트워크는 인간의 두뇌에 존재하는 뉴런을 가상의 형태로 세팅하고 각각의 뉴런에 임의의 가중치를 부여하고(사전에 뉴런이 가중치를 조정하면서 반응하도록 프로그래밍), 연결하는 방식이다. 또한 대량의 데이터를 입력하고 하나의 신경네트워크가 하나의 대상을 포착하도록 훈련시키며 사후적으로 가중치를 조정하고 대상 식별 정확도를 높이는 작업을 반복한다.

이러한 신경네트워크를 활용해 발전시킨 알고리즘이 심층 학습이다. 심층 학습은 수십억 개의 뉴런을 복수의 계층 구조로 연결하여 순방향으로 연산하고, 예측과 정답 사이의 오차를 이전 계층으로 다시 보내주면서 학습한다. 기계는 정해진 신경망 구조를 통해 마치 인간이 새로운 것을 학습하는 것처럼 경험을 중심으로, 예측과 정답 사이의 오차를 계속해서 학습하며 알고리즘을 형성한다.

예를 들어, 동물 분류 시스템의 경우 1) 첫번째 레이어에서 강아지 이미지의 밝기, 색채 등의 대략적 분포를 학습, 2) 두번째 레이어에서 강아지의 형태, 그림자 같은 더 복잡한 특성을 식별, 3) 세번째 레이어에서 눈, 귀, 코와 같은 구체적 형상을 훈련, 4) 마지막 레이어에서 모든 것을 총체적으로 결합하여 학습 값을 출력한다. 마지막에서 출력된 학습 정보와 입력 값의 상관관계를 찾고, 결과값이 신뢰할 만한 수준까지 학습을 반복한다.

그 결과 심층 학습이 된 동물 분류 시스템에선 기계 학습 과정처럼 인간이 강아지의 특성을 모두 분류 해주지 않아도 기계가 정해진 신경망을 통해 스스로 강아지 데이터를 학습해 강아지의 특성을 학습하고, 분류할 수 있다. 심층 학습은 기계가 스스로 학습한다는 차원에서 기계 학습의 일부지만, 기계가 스스로 데이터를 분류하고 학습할 수 있다는 점에서 더 세분화된 개념이다.

## 3) 대규모 언어 모델(LLM)

- 대규모 언어 모델은 방대한 양의 데이터를 학습하여 자연어 처리 작업을 수행하는 모델이다. 대규모 언어 모델은 주로 '트랜스포머 아키텍처'를 기반으로 한 심층 학습 모델을 사용하는데, 그 이유는 트랜스포머 아키텍처가 가진 주요 특징인 Self-attention 메커니즘에 있다.
- 트랜스포머 아키텍처는 입력된 정보 값을 토큰화하고 Self-attention 메커니즘을 통해 토큰 값들 간 관계를 파악한다. 특정 토큰의 본래 의미에 가중치를 두어 집중(Attention)하되, 나머지 토큰들과 관계를 참고해 맥락을 파악한다. 각 토큰은 다른 모든 토큰과 관계에 점수를 매기고, '관심 점수'로 어떤 값이 출력되어야 하는지 결정된다.
- 예를 들어, '우리 강아지는 벽지도 다 뜯고, 참 똑똑해.'라는 문장을 학습할 때, 본래의 '강아지', '벽지', '뜯고', '참', '똑똑해' 각 단어의 본래 뜻에 집중하되, 단어 간의 관계에 관심 점수를 매겨 맥락을 파악해 숨은 의도를 알아낸다. 트랜스포머 아키텍처는 대규모 언어 모델이 심층 학습을 할 때 데이터의 맥락을 파악하는 것에 이점이 있다.

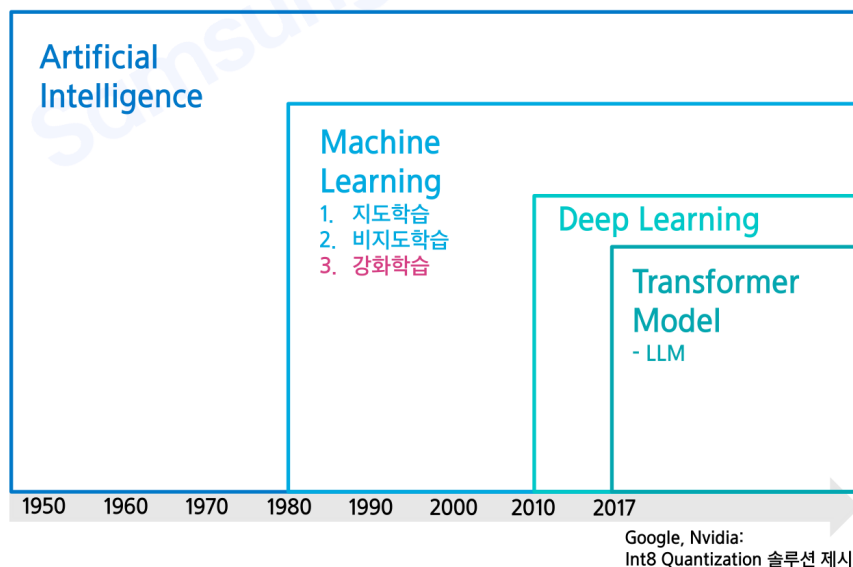
## 4) 강화 학습 (Reinforcement Learning)

- 기계 학습의 학습방법은 1) 지도학습 2) 비지도학습 3) 강화 학습 총 3가지로 분류된다. 지도학습은 정답이 있는 데이터를 분류, 회귀 방식으로 학습시키는 방식이고, 비지도학습은 정답이 없는 데이터를 군집화하여 새로운 데이터에 대한 결과를 예측하는 방식이다.
- 강화 학습은 지도학습, 비지도학습과는 조금 다른 개념으로, '보상'이라는 개념이 존재한다. 강화 학습은 어떤 환경에서 '주체'가 주어진 '환경' 속에서 현재의 '상태'를 고려하여 취할 수 있는 '행동' 중 가장 높은 '보상'이 무엇인지 학습하는 방식이다. 언뜻 지도학습과 비슷한 개념 같지만 명확한 정답이 존재하는 것이 아니라, '주어진 환경에서 행동해서 얻은 보상을 통해 학습한다는 점'에서 차이가 있다. 최고의 보상을 얻을 수 있도록 수많은 행동 반복을 통해 알고리즘을 강화한다는 점에서 강화 학습이라고 불린다.

### 5) 양자화 (Quantization)

- 양자화는 간단히 말해 모델을 더 작은 비트로 표현해서 경량화하는 기법을 의미한다.
- 앞서 심층 학습은 각 뉴런을 가상의 형태로 세팅하고, 각 뉴런에 가중치를 부여하는 신경망 구조로 되어있다고 설명하였다. 각 뉴런의 활성화 함수(뉴런을 가상의 형태로 세팅한 방식)와 가중치는 32비트의 부동소수점(이하 FP32) 데이터 타입으로 구현되어 있다. FP32로 모델을 표현한다면 더 세부적인 숫자를 표현할 수 있는 만큼, 모델의 정확도가 올라간다.
- 그러나 모바일 기기나 엣지 디바이스와 같이 리소스가 제한된 환경에서 FP32로 모델을 표현하기엔 메모리, 프로세서 성능, 연산 비용 등의 한계가 있다. 따라서 모델을 경량화해 연산 속도를 높이고, 메모리 사용량을 줄여 계산 비용을 줄이려는 시도가 나타나고 있다.
- 경량화 연구는 크게 1) 알고리즘 자체를 효율적으로 설계, 2) 모델의 파라미터를 줄이거나 압축하는 연구로 나뉘는데, 양자화는 모델의 파라미터를 압축하는 방식에 속한다. 양자화는 32비트 부동소수점으로 저장된 모델의 파라미터를 16비트/8비트/4비트 등 정수로 맵핑하여 모델의 크기를 줄인다. 일반적으로 8비트 정수로 맵핑하는 방식이 널리 쓰이는데, 구글과 엔비디아가 각각 연구한 결과에 따르면, 8비트 정수로도 모델의 파라미터를 충분히 표현할 수 있다고 한다. 8비트 정수로 모델을 표현하면 32비트 부동소수점 데이터 타입에 비해 모델 크기 및 메모리 대역폭이 4배 감소할 수 있다.
- 모델을 경량화하여 연산비용을 줄일 수 있다는 이점이 있지만, 양자화도 단점이 있다. 32비트 부동소수점에서 8비트 정수로 채널이 줄어들며 정보가 손실되기 때문이다. 예를 들어 32비트 부동소수점에서 2.13과 같이 분류하던 것을 8비트 정수로 2로만 분류할 수 있다. 정확한 실수가 아닌 그 근사치 정수를 활용하기 때문에 기존 모델에 비해 정확도가 떨어질 수밖에 없다. 따라서 이러한 한계를 극복하기 위해 학습 후 양자화, 양자화 인식 학습 등의 연구도 현재 진행되고 있다.

### 인공지능 기술 발전 개념도



자료: 삼성증권 정리

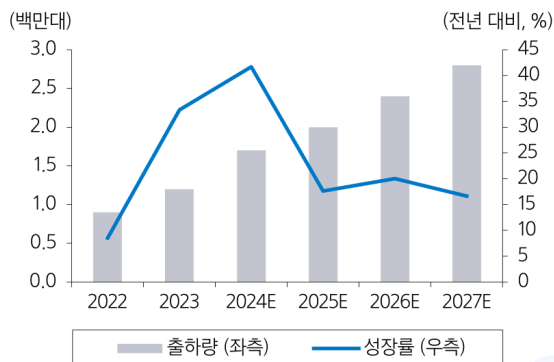
## 부록 2: 인공지능 밸류체인

생성 AI의 밸류체인은 AI 서비스를 개발하기 위한 도구인 1) AI 인프라, 2) 클라우드와 파운데이션 모델, 그리고 3) AI 개발 레이스의 종착지인 최종 서비스/애플리케이션 세 가지 영역으로 구분할 수 있다.

## 인공지능의 시작, 인프라

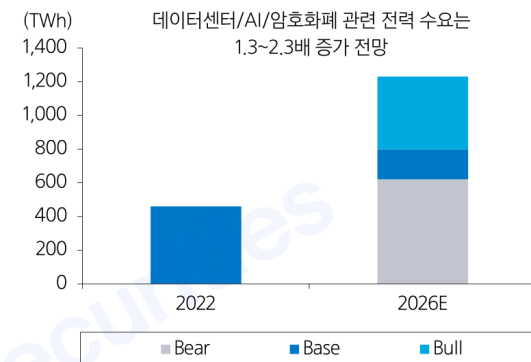
AI 서비스를 개발하기 위해서는 데이터센터 수주의 막대한 컴퓨팅 파워가 필요하다. 막대한 컴퓨팅 파워로 중무장한 AI 서버, 그리고 이들 서버 간 고속 연결을 지원해주는 네트워킹 장비 등이 필요하다.

AI 서버 출하량 추이 및 전망



자료: 언론 보도

데이터센터/AI/암호화폐 관련 전력 수요 전망



자료: IEA

**AI 가속기와 고대역 메모리 (HBM).** 2022년 말 ChatGPT의 출현 이후, AI 개발 경쟁이 본격화되며 AI 서버 출하량이 급증하고 있다. 우리가 흔히 생각하는 일반 서버와 단순 성능 차이를 넘어 한 가지 차이점이 있다면 바로 AI 가속기의 탑재 여부다. 일반 서버는 CPU와 메모리로 구성된다고 할 수 있다. 그러나 CPU는 AI 개발에 있어서 구조적으로 불리하다.

## 일반 서버 BOM breakdown

(달러)	수량	ASP	전체 비용	비고
CPU (개)	1	2,600	2,600	Intel Xeon Gold 32코어 (Ice Lake)
DRAM (GB)	550	1.6	880	DDR 4 64GB 모듈 기준
SSD (TB)	4	307.2	1,260	1GB=0.3달러
기타	n/a	n/a	1,960	
서버 가격			6,700	

자료: 삼성증권

## AI 서버 BOM breakdown (NVIDIA DGX H100)

(달러)	수량	ASP	전체 비용	비고
CPU (개)	2	10,000	20,000	Intel Xeon Platinum 32코어 (Sapphire Rapids)
GPU (개)	8	25,000	200,000	NVIDIA H100
HBM (GB)	640	9.6	6,144	HBM3
DRAM (GB)	2,048	7.8	16,000	DDR 5 128GB 모듈 기준
SSD (TB)	31	307.2	9,437	1GB=0.3달러
기타	n/a	n/a	4,563	
서버 가격			250,000	일반 서버 37개 구매 비용 수준

자료: 삼성증권

**AI 가속기의 역할.** AI 서비스를 개발하는 과정(AI 훈련)을 비약하자면 방대한 양의 데이터를 처리하는 것이다. 문제는 CPU는 연산 속도는 빠르지만 적은 수의 코어로 구성되어 있고 무엇보다 '순차 처리' 혹은 '직렬 처리' 방식이라는 점이다. 즉, 각 코어가 데이터를 하나씩 순차적으로 처리하는 구조다.

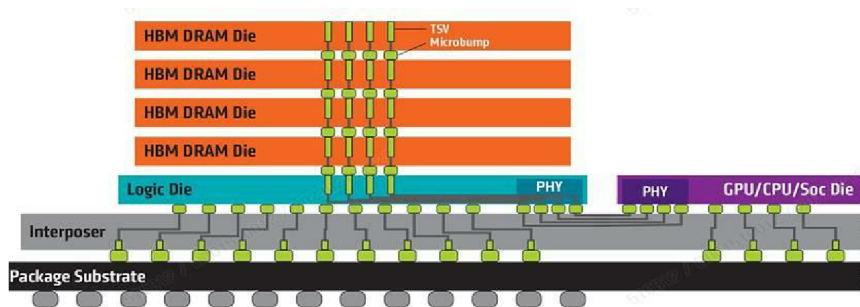
때문에 CPU의 옆에서 특정 연산을 대신 처리해줌으로써 전체 데이터 처리 속도를 단축시켜 주는 반도체들이 주목 받게 되었다. 이를 흔히들 가속기(accelerator)라고 부르는데, 엔비디아의 GPU가 대표적이다. GPU는 4~300개 정도의 코어로 구성된 CPU와 달리 수천, 수만개의 코어로 구성되어 있고, 각 코어가 한 번에 여러 가지 연산을 할 수 있는 '병렬 처리' 방식이다. AI 훈련에 있어서 구조적으로 더 효율적일 수밖에 없는 것이다.

일반 서버와 AI 서버의 가격 차이는 30배 넘게 나는데, 가격 차이의 대부분은 가속기에서 발생하게 된다. AI 훈련의 가장 핵심이 되는 반도체이다 보니 기업들은 기꺼이 개당 3만 달러 이상을 지불하고 있다. AI 최전선에 있는 마이크로소프트, 구글 등의 hyperscaler들은 1분기 실적 발표에서도 여기서 투자를 더 늘릴 의지를 표명한 바 있다. 이에 더해 각자가 자신의 시스템에 최적화된 자체 AI 반도체 설계에 나서고 있을 정도다.

**GPU를 사용하기 위한 핵심 자원, HBM.** AI 서버는 한 가지 차이점이 더 존재하는데 바로 HBM(High Bandwidth Memory)의 차이 유무다. 데이터 처리 속도를 높이기 위해서는 가속기의 연산 능력도 중요하지만, 메모리 성능이 반쳐주지 않는다면 병목이 발생할 수밖에 없다. 즉, 메모리와 프로세서 간의 데이터 전송 속도 역시 개선이 요구된다. 시장은 이를 해결하기 위해 더 큰 대역폭의 솔루션을 개발해왔고 이의 결과물이 HBM이다. 물론 이는 엄밀히 말해 서버에 별도로 내장되는 것이 아닌 AI 가속기에 패키징을 통해 연결되는 구조다.

HBM이 인공지능에 반드시 필요할 것이라고 보는 시각은 많지 않았다. 앞서 설명하였듯이 인터넷 초기에도 Rambus 디램이라는 고대역 메모리를 시도하였고, 결국 높은 라이선스 비용 등을 거부한 제조업체들이 DDR이라는 새로운 표준을 지원하였다. 이처럼 혁신도 가능성이 중요하다. HBM은 직전 고대역 메모리였던 그래픽 디램 대비 많이 비싸다. 작년에는 일반 범용 디램 대비 6~7배가 비싸게 거래되었다. 생산 수율도 낮다. 그럼에도 불구하고 엔비디아를 비롯한 고객들이 HBM을 우선하는 것은 많은 데이터 연산이 필수적인 인공지능에는 비용보다 성능이 우선시된다.

#### HBM을 도입한 그래픽 카드 구조도



자료: AMD

## 클라우드, 인프라와 서비스의 연결 다리

## 클라우드 서비스(IaaS)

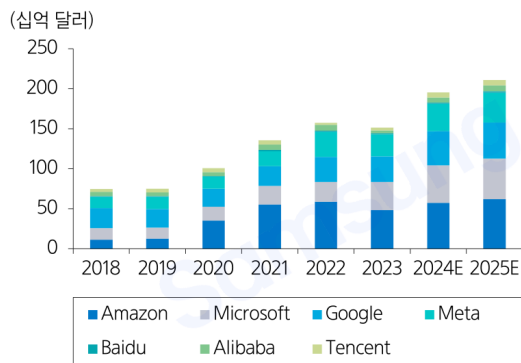
AI 서버만 있으면 AI 서비스를 개발할 수 있을까? 그럴 수도 있지만 문제는 일단 AI 인프라를 구축하는 것 자체가 자본금을 고려하면 굉장히 큰 모험이다.

예를 들어, 연초 메타 플랫폼스의 저커버그는 올해 엔비디아 H100 GPU를 35만개 구입 예정이라고 언급한 바 있다. 대량 구매이니 개당 단가를 25,000달러라고 소폭 낮게 가정해보자. 그럼에도 순수 GPU 구매 비용만 87.5억 달러에 육박한다. 우리나라 돈으로 11조원에 달하는 금액이다. 막대한 자본 투입이 가능한 극소수들만 AI 인프라를 꾸릴 수 있는 것이다.

게다가 구매하고 싶어도 작년에는 공급 부족으로 인해 대형 고객사들에 우선적으로 물량이 배분되어 왔다. 일반 기업들로서는 사실상 AI 인프라를 직접 꾸리는 것이 비현실적이고, 결국 클라우드 업체들의 인프라를 대여할 수밖에 없게 된 것이다.

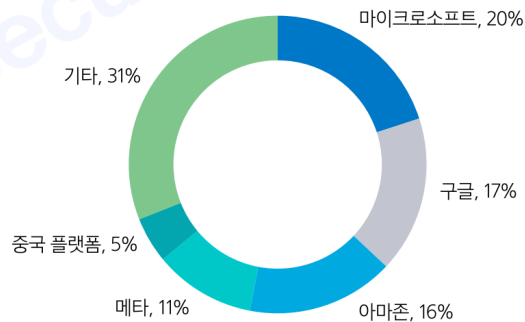
밸류 체인 관점에서 보면 클라우드 업체들은 AI 인프라와 서비스의 연결 다리라고 할 수 있을 것이다. 그리고 이미 인프라 관련 업체들 다음으로 수익을 창출하고 있다. AI 개발 레이스가 본격화된 이후, 작년 하반기부터 클라우드 업체들의 매출 성장률이 다시 가속화되기 시작했다. 마이크로소프트의 경우, 3월 분기 Azure 매출 성장률 31% 중 7%pt가 AI에 기인했다고 언급할 정도다.

상위 7개 hyperscaler CAPEX 추이 및 전망



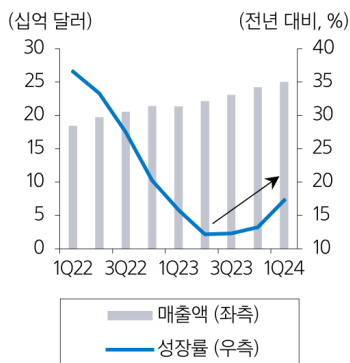
자료: 각 사, FactSet

고사양 AI 서버 수요 breakdown



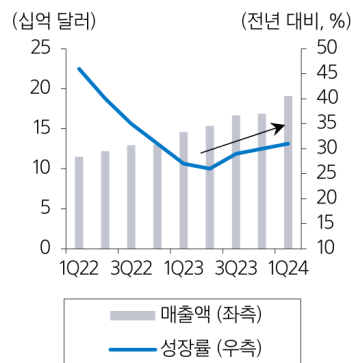
자료: TrendForce

AWS: 클라우드 매출액과 성장률 추이



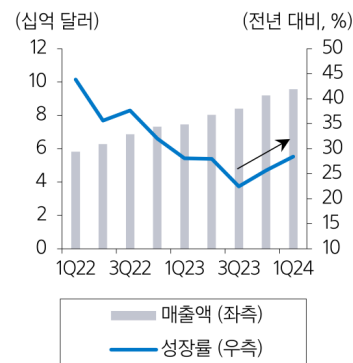
참고: Calendar year 기준 / 자료: Amazon

Azure: 클라우드 매출액과 성장률 추이



참고: Calendar year 기준 / 자료: Microsoft

GCP: 클라우드 매출액과 성장률 추이



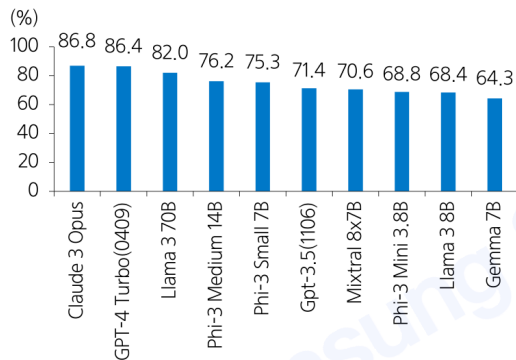
참고: Calendar year 기준 / 자료: Alphabet



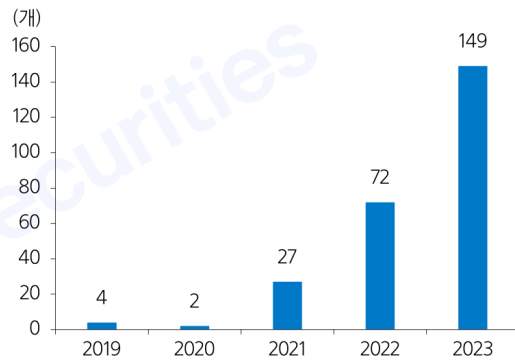
**파운데이션(Foundation 모델)**

파운데이션 모델은 생성 AI 서비스의 근간이며, 또한 클라우드 업체들이 AI 관련하여 수익을 창출하는 또다른 분야다. 아마존이 1분기 실적에서 언급한 것처럼 수많은 기업이 모델 개발을 진행하고 있다. 파운데이션 모델은 클라우드 기반의 API 형태로 제공된다. 기업 고객 및 개발자는 기학습된 모델을 활용해 자체 보유 데이터로 추가 학습하거나, 입맛에 맞게 Fine Tuning(파인 튜닝)을 진행한다. 최적의 생성 결과를 얻기 위해서 복수 모델을 동시에 활용하기도 한다.

산업 발전 초기 단계에서는 생성 AI 서비스 시작점 챗GPT의 기반이 되는 언어 모델이 주목 받았다. 현재 언어 모델 내 최선두 모델(SOTA, State-Of-The-Art) 자리는 출시 1년이 지났지만 오픈AI GPT-4(23년 3월 출시)가 차지하고 있다. GPT-4 등장 이후 파운데이션 모델은 극한의 성능을 추구하는 프론티어(Frontier) 모델과 파라미터 개수와 사이즈를 적당히 줄여 다양한 디바이스 적용과 손쉬운 커스터마이징 및 서비스 활용을 목표로 하는 소형모델 sLLM 개발로 분화되어 발전했다. 다만 최근에는 프론티어 모델과 소형 모델 간 경계도 흐려지고 있다.

**벤치마크 상 경계가 흐려진 프론티어 모델과 소형 모델**

자료: 삼성증권 정리

**연간 파운데이션 모델 추이**

또한, 오픈 소스 모델의 존재는 다양한 파생 모델 개발로 이어지며 생태계를 확장시키고 있다. 분화되었던 개발 트렌드의 혼합과 오픈 소스 모델 등장에 따라 파운데이션 모델의 발전 속도는 더욱 빨라지고 있다. 23년 연간 발표된 파운데이션 모델은 149개로 전년 72개 대비 2배 수준으로 상승했다. 24년에도 이어지는 파운데이션 모델 출시 러시를 고려하면 수치는 한번 더 도약할 것으로 전망한다.

GPT-4 출시 후 1년이 넘는 시간이 지났고, Gemini Ultra(구글), Claude 3 Opus(앤스로픽), Mistral Large(미스트랄 AI), Inflection 2.5(인플렉션 AI) 등 경쟁 기업의 AI 모델은 GPT 4와 유사하거나 이를 뛰어넘는 성능을 보여주고 있다. GPT 4는 이제 당연하게 도달해야 할 기준이 되었다. 경쟁 모델도 단순 벤치마크 테스트 결과뿐 아니라 훨씬 거대한 Context Window를 제공하거나, 적은 컴퓨팅 자원을 사용하거나, 다양한 언어에 최적화되어 있는 등의 차별점을 강조하고 있다. 인공지능이 생성할 수 있는 것은 단순한 문장에 국한되지 않는다. 이미지, 동영상, 3D 콘텐츠 등에 초점을 맞춘 다양한 생성 AI 모델이 출시되고 있다.

결국 대중이 궁극적으로 원하는 것은 여러 형태의 생성 AI 파운데이션 모델이 결합된 멀티모달(Multi Modal) 모델이다. 일부 프론티어 언어 모델에서 멀티모달 기능을 지원하고 있지만 완벽한 수준은 아니다. 당연히게도 오픈AI GPT 시리즈의 다음 세대(GPT 4.5 또는 GPT 5)에 대한 이목이 집중되고 있다.

샘 알트먼은 GPT 차세대 모델이 모든 측면에서 이전 모델보다 더 나을 것이라고 코멘트했다. 추론, 코딩 등 능력 외에 오디오와 글쓰기가 훨씬 개선되었으며, 신규 모델은 더 똑똑하고 더 빠르며 멀티 모달이라는 점도 강조했다. 동반되고 있는 로봇 산업 및 자율 에이전트 산업 발달을 고려하면 차세대 AI 모델 혁신과 연계된 시너지 효과가 더욱 기대되는 시점이다.

### 인공지능의 완성, 서비스

서비스 부문은 생성 AI 기술이 실제 프로덕트로 구현되는 종착 단계다. 스타트업을 포함해 다양한 플랫폼 및 SW 기업이 포진하고 있다. 챗GPT 같은 AI 챗 봇도 서비스 예시 중 하나다. '생성'에 초점을 맞추면 챗 봇의 서비스 적용 범위도 요약, 번역, 코드 작성 등으로 확장이 가능하다. 또한 이미지 및 동영상 생성은 콘텐츠 분야 혁신을 불러올 수 있다.

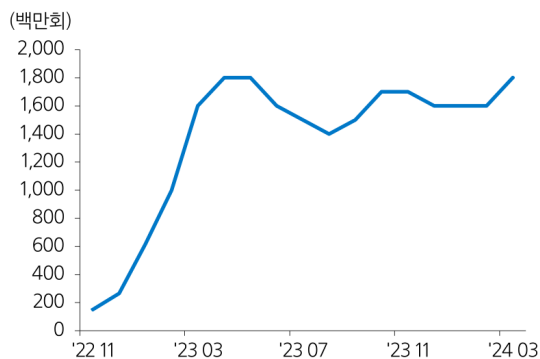
서비스 분야에서 고려해야 할 첫번째 이슈는 수익화다. 생성 AI 서비스의 수익화 전략은 1) 유료 구독 모델과 2) 크레딧 정책으로 분류할 수 있다. 구독 모델은 세부적으로 차등 플랜을 적용하거나 기존 프로덕트 접목 후 가격을 인상하는 방법 등이 활용된다. 크레딧 정책은 건 당으로 소비되는 형태가 일반적으로 쿼리를 처리하기 위해 소요되는 컴퓨팅 비용을 충당하는 개념이다. 초기 유저 확대를 위한 무료 전략도 존재하나, 서비스 활용에 따른 비용이 무한하게 증가할 수 있다는 점을 고려하면 지속성 가능성은 높지 않다.

서비스 수익화 확인 타임라인은 지연되고 있다. 대표적으로 어도비는 이미지 생성 Firefly의 편집툴 적용을 통해 23년 말부터 일정 수준의 매출 상승 효과를 기대했으나, 실적 확인 기대감은 올해 하반기로 밀린 상황이다. 다만 시차가 있을 뿐 수익화 자체는 충분히 달성 가능하다고 판단한다. 마이크로소프트 코파일럿 유저의 고성장과 서비스나우의 생성 AI 기능이 포함된 Pro + SKU의 높은 침투율은 긍정적 전망의 기반이다.

수요에도 불확실성이 존재한다. AI를 통한 생산성 극대화가 가능하다는 점에서 높은 수요를 기대하는 시각이 많지만, 챗GPT 트래픽(월 방문)은 박스권 움직임을 보이고 있다. 경쟁 챗 봇이 등장한 영향도 고려해야 하지만, 할루시네이션(가짜를 진짜처럼 생성) 이슈를 비롯해 완벽한 서비스가 아니라는 점이 걸림돌이다. 업무 처리 속도를 개선해주지만 가끔 부정확한 답변을 내놓는다면 제한적 수요에 그칠 수 있다. 이러한 이슈를 돌파하는 방법은 AI 모델 성능 개선과 고도화다.

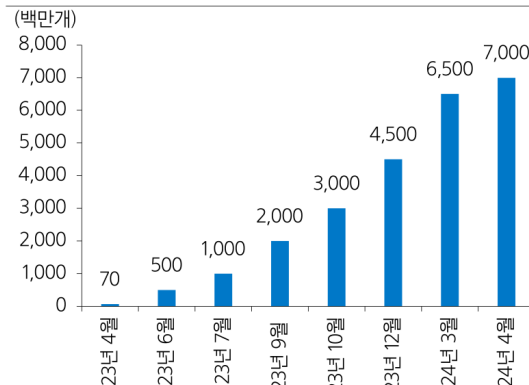
아마존은 실적 컨퍼런스 콜에서 의미 있는 내용을 언급했다. 생성 AI 어플리케이션 증가에 따라 운영 환경에 배치되는 시점에는 상당한 비용이 소요될 것으로 전망하며, 학습은 주기적이지만, 추론은 항상 일어난다는 측면에서 추론 지출이 더 클 것으로 전망했다. 생성 AI 밸류체인은 상호 연결되어 있다. 생성 AI 서비스의 증가는 추론 수요의 성장으로 이어진다. 클라우드와 반도체 인프라의 새로운 트렌드로 이어질 수 있다.

제한적 성장을 하는 챗GPT 월 방문 추이



자료: Similarweb

아쉬운 수익화와 달리 증가하는 어도비 Firefly 이미지 생성량



자료: Adobe

## 부록 2-1: 미디어·엔터테인먼트 - 비용은 낮추고, 퀄리티는 높이고

## Potential &amp; Benefit

**딥러닝으로 더욱 정교해진 효과.** 콘텐츠 제작 업계에서는 AI를 활용한 딥페이크 (Deepfake), CG (Computer Graphic) 등이 화두다. 특정 인물의 얼굴, 신체 부위 등을 영상에 합성하는 것을 의미하는 딥페이크는 AI 기술의 발전과 함께 정교해지고 있다. 과거에는 사진, 영상의 합성이 어색해 쉽게 진위를 가려낼 수 있었다. 그러나 기존 사진, 영상을 통해 얼굴, 특정 부위 등을 충분히 학습한 이후 다른 동영상에 합성하는 방식의 딥러닝 (Deep Learnings)을 도입하며 점점 더 실재와 구분하기 어려워지고 있다.

**작품의 비용은 낮추고, 퀄리티는 높이고.** 미디어·콘텐츠 하나를 창작하는데는 상당히 많은 인력과 자본이 투입된다. 작품 기획 단계부터 배우와 제작 스태프 및 촬영 장소 섭외, CG 작업 등에 이르기까지 많은 시간과 다수의 노력이 들어간다. 콘텐츠 제작 업계에 AI 기술이 도입되면서 제작 단가를 낮추고 제작 시간을 단축시켜 콘텐츠 제작에 대한 문턱을 낮추는 한편 작품의 퀄리티를 높이는 방향으로 변화가 나타나고 있다.

**획기적인 비용 감소.** 올해 공개된 한국 드라마에서도 AI 딥페이크를 활용한 장면이 등장했다. 넷플릭스 오리지널 '살인자o난감'에서 주연 배우 손석구의 아역 배우가 높은 싱크로율을 보여 화제가 됐는데, 딥페이크 기술로 구현한 결과물이다. Statista의 조사 결과에 따르면 현재 미국 TV & 영화 산업에서 AI 기술을 가장 많이 활용하고 있는 것은 바로 '특수 효과'다. 할리우드 업계는 디지털 트윈(Digital Twin: 현실에 존재하는 장비·기계 등 사물들을 실재와 동일한 3차원 모델을 통해 가상세계에 구현하는 기술) 기법을 활용해 대역 배우를 만들거나, 배경에 엑스트라를 생성하는 방식으로 AI 기술을 적용하고 있다. 특수효과를 많이 사용하는 할리우드 블록버스터 영화들은 영상을 촬영하는 시간보다 리얼리티를 높이기 위한 후반 작업에 더 많은 시간이 소요되는 경우도 있다. 적합한 배우를 찾기 위해, 보다 질 높은 콘텐츠를 만들기 위해 공을 들여야 했던 시간과 비용, 노력 등이 AI 기술을 통해 단축되면서 제작 효율성을 높이는 한편 콘텐츠의 퀄리티를 향상시키고 있다.

과거 드라마 '주몽'의 전쟁 장면



자료: KBS, 삼성증권

AI가 도입된 현재 드라마 '고려러란전쟁'의 전쟁 장면



자료: 비브스튜디오, KBS, 삼성증권

**제작과 마케팅에도 적용.** 특수 효과를 넘어 할리우드 업계는 초고 등 대본 작성, 홍보·마케팅·프로모션 등에도 AI 기술 활용을 확대하고 있는 추세다. 넷플릭스, 디즈니, 아마존 등 글로벌 OTT 서비스를 영위하고 있는 대형 미디어사들은 현재 제작과 마케팅 측면에서 AI 기술을 활용하고 있다.

**창작에도 인공지능.** 엔터테인먼트 업계에도 버추얼 아티스트, 음악 생성 AI 등이 속속 늘어나며 변화의 바람을 몰고 있다. 올해 초 CJ ENM이 투자한 AI 스타트업 '포자랩스'가 개발한 AI 기반의 작사·작곡·가창 프로그램인 LAIVE(라이브)는 템플릿, 장르, 무드, 보컬리스트 등 몇 번의 클릭과 원하는 가사 키워드 입력으로 10분 만에 짧은 음악을 완성했다. 이미 CJ ENM의 예능에서 배경 음악으로 활용하고 있다.

### Barriers to overcome

**스토리가 재미있어야 명작이지.** AI 기술이 기획, 제작, 광고, 마케팅 등 미디어·엔터테인먼트 산업 환경을 바꾸고 있는 것은 사실이다. 그러나 짚고 넘어가야 할 부분은 AI 기술을 활용한 콘텐츠 제작이 분명 영향력이 커지기 시작했지만 기술과 접목된 영역을 제외한 즉 지식 재산권(IP)과 관련 있는 '크리에이티브' 영역은 아직까지는 기술의 영역만큼 빠른 속도로 시장을 잠식해 나가고 있지는 않다는 점이다. 풀어야 할 현실적 제약이 남아있기 때문이다.

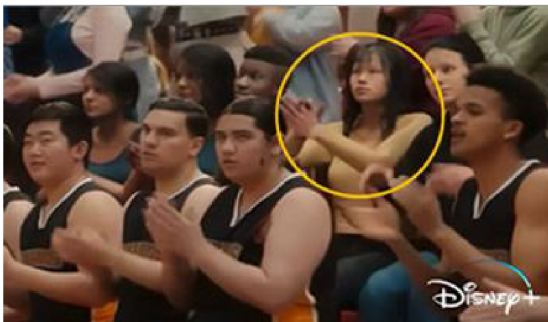
**아주 불편한 불쾌한 골짜기.** 콘텐츠 제작 밸류체인 상에서 생성 AI를 활용하는 영역이 늘어나고 있지만 아직 초기 단계에 불과해 기술적 완성도가 낮고 오류 발생 가능성이 높아 안정적인 콘텐츠 제작은 어렵다. AI기술이 급격히 발전하면서 몇 년 전과 비교할 때 AI를 활용한 콘텐츠가 어색함과 부자연스러움에서 오는 '불쾌한 골짜기'를 넘어서기 시작했으나 완전히 기존 영역을 대체했다고 말하긴 힘들다.

기본적으로 생성 AI는 기존에 출판되었거나 공개된 영상, 텍스트, 뉴스 등의 콘텐츠를 학습 데이터로 사용해 훈련을 받았다. 정보가 넘쳐나는 시대인 만큼 올바른 데이터와 잘못된 데이터를 식별할 수 있는 변별력이 필요하다. 또한 잘못됐거나 틀린 내용이 아니더라도, 불쾌감을 일으킬 수 있는 표현이나 장면 등이 제작물에 포함되지 않도록 판단할 수 있는 능력도 필요하다. 만약 AI가 학습한 데이터가 차별적이고 편향적인 성향을 갖고 있었다면 인종·성차별적 콘텐츠 등 윤리적 문제의 소지가 있는 제작물이 나올 수 있기 때문에 AI기술을 활용해 작업을 하더라도 현재는 사람이 최종적으로 검토를 하고 책임을 지는 경우가 대다수다. 한국 콘텐츠 제작사에게 AI 활용도에 대해 조사해본 결과 '기술적인 영역'은 일부 커버할 수 있는데 아직까지는 '감성적인 부분'은 해결해야 할 과제가 있는 것 같다는 답을 얻었다. 특수효과 작업을 할 때도 AI를 기반으로 하면서 시간적, 비용적 측면에서 분명히 절감이 나타나고 있지만 여전히 제작진이 최종 확인이 필요한 상황이라고 보고 있었다.

**저작권, 초상권 침해 이슈.** 생성 AI를 활용한 작품들이 제작되면서 전 세계적으로 미디어·엔터테인먼트 업계의 화두가 된 것은 '저작권' 및 '초상권' 이슈이다. 작년 글로벌 영화산업의 중심지인 미국 할리우드의 배우와 작가가 63년 만에 동시에 파업에 나서며 할리우드가 멈춰 섰다. 파업의 원인은 바로 'AI'였다. 영상 콘텐츠를 제작하는데 작가의 대본이 필요하고, 이에 맞춰 연기할 배우가 필요한데 AI가 만 들어내는 대본으로 인해 작가들이 곤욕을 치르며 저작권, 창작권 등이 이슈가 됐고, 배우들의 초상권 침해가 우려됐기 때문이다.

실제로 지식 재산권(IP)의 저작권 침해와 표절 문제 등이 불거지고 있고, 어디까지가 저작권의 대상인지도 아직까지 명확하게 정의되지 않았다. 창작물에 대해 AI가 얼마나 관여했는지, 표절 여부를 판별하는 기술, 기준이 중요해질 것으로 예상된다. 아직까지 미디어·엔터테인먼트 업계에서 뜨거운 감자와도 같은 저작권, 초상권 등의 민감한 이슈가 확실히 해결되지 않은 만큼 밸류체인 상 관련 영역에서는 기존과 거의 유사한 방식으로 제작 환경이 돌아가고 있다.

### 영화 '프롬 팩트'에 등장한 AI 엑스트라, 아직은 어색한 모습



참고: 23년 3월 디즈니 채널에서 공개된 영화 '프롬 팩트(Prom Pact)'의 장면에 AI 엑스트라가 활용  
자료: Disney, 삼성증권

### 배우 톰 행크스가 SNS에 올린 사진- 초상권 침해 관련



참고: AI 기술을 활용해 딥페이크로 만들어진 이미지가 치과 보험을 홍보하는 영상이 있다며 주의하라고 SNS(인스타그램)에 업로드한 사진  
자료: 인스타그램, 삼성증권

## 부록 2-2: 금융 - 디지털 금융의 고도화 - 매출은 키우고 비용은 줄이고

## Potential &amp; Benefit

금융 산업은 이미 AI를 적극적으로 업무에 접목시켜 왔다. 일례로, 한국신용정보원이 발간한 '금융 AI 시장 전망과 활용 현황: 은행권을 중심으로' 보고서에 따르면, 1) 21년 기준 전 세계 인공지능 시장 규모 581억 달러 중 금융은 113억 달러로 19.5%를 차지하고 있으며, 2) 국내에서도 같은 기간 전체 시장 규모 3.2조원 중 금융은 0.6조원, 18.8%를 차지하고 있다.

## 금융이 인공지능을 활용하는 4가지 유형: 매출은 키우고 비용은 줄이고

- 방대한 데이터 처리를 통한 신규 수익원 창출
- 소비자에게 개인화 서비스 제공을 통한 가치 제고
- 업무 프로세스 자동화 등을 통한 운용 비용 절감
- 복잡한 금융규제 및 컴플라이언스 요구 사항에 대한 효과적 대응

## 금융산업의 AI 접목 기대 효과 1) 업무 자동화 → 비용 효율성 제고 → 소비자 만족 증가

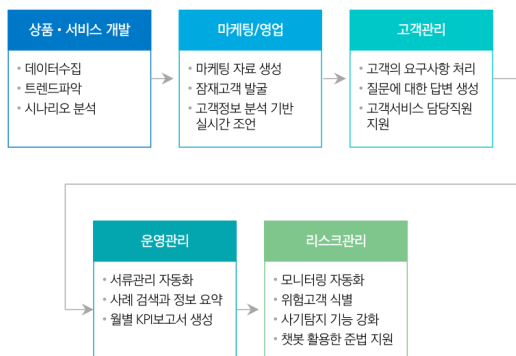
금융사들은 생성형 AI 출현 이전부터 AI 접목을 통해 업무 자동화를 통한 비용 효율성 제고를 추구해 왔다. AI와 빅데이터를 접목시키고, 이를 업무 프로세스에 적극 활용함으로써 1) 투입되는 노동력을 절감하는 것은 물론, 2) 기존에 비해 소요되는 업무 시간과 3) 업무상 발생 가능한 오류를 대폭 줄임으로써 고객 만족도까지 높이는 것이 주된 목적이다.

특히, 이 과정에서 초점을 뒀다 할 부분은 단순한 인력의 축소가 아닌, 기존 인력의 수익성 향상이다. 단순 업무의 자동화 등을 통해 기존 인력을 보다 부가가치가 높은 분야에 투입, 서비스의 질과 수익성을 높일 수 있기 때문이다. 가령, 대출 심사의 경우, 과거에는 기존 인력이 단순 업무 처리를 위해 많은 시간을 할애하거나, 이를 막기 위해 외주를 썼었으나, 업무 자동화는 관련 비용을 절감하거나, 기존 인력을 보다 고부가가치 업무에 배치함으로써 수익성을 높일 수 있을 것이다.

## 금융산업의 AI 접목 기대 효과 2) 생성형 AI 접목 → 대고객 서비스 강화

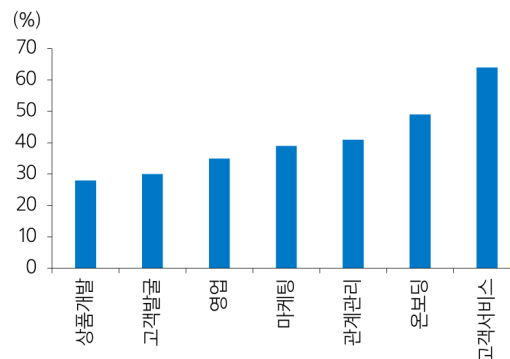
우리금융연구소가 지난 2월 발표한 바에 따르면, 생성형 AI는 상품·서비스 개발, 마케팅 및 영업, 고객 관리, 운영관리, 리스크 관리 등 전 부문에 걸쳐 적용 가능하다. 그중에서도 우리가 주목하는 부분은 생성형 AI의 질의 응답 및 콘텐츠 편집, 제공 등의 기능을 활용한 대고객 서비스 강화에 있다. 실제로, EY가 글로벌 151개 은행을 대상으로 조사한 결과에 의하면 생성형 AI로 프론트오피스 업무에서 비용 절감과 생산성 향상이 기대되는 업무 중 가장 높은 응답률을 보인 분야는 고객 서비스로 나타났다.

## 생성형 AI 은행업 적용 범위



자료: 우리금융경영연구소, 삼성증권

## 생성형 AI 적용시 비용 절감 및 생산성 향상 기대 업무



자료: EY, 우리금융경영연구소



**Barriers to overcome****금융의 생성형 AI 접목의 한계: AI를 신뢰할 수 있는가?**

다만, 현실점에서 한계 또한 명확하다. 금융업에서의 상품 판매에 가장 중요한 것 중 하나는 고객에 대한 설명 책임이다. 하지만, 생성형 AI는 스스로 알고리즘을 생성하고 이를 수시로 바꾸거나 진화시키는 만큼, 생성형 AI가 추천하는 결과물의 이유나 과정에 대해 설명할 수 없다는 문제가 발생한다. 그렇다면 고객 신뢰를 얻기에는 한계가 있다.

만약, 생성형 AI가 제시한 상품 혹은 결정이 가져올 리스크를 사전에 인지하기 어려운 것은 물론, 고객의 손실로 이어졌을 때, 이에 대한 사후 관리도 어려울 수 있으며, 이는 궁극적으로 금융사의 신뢰도를 훼손시키는 결과로 이어질 수 있다. 실제로, 23년 BofA의 CEO는 실적 컨퍼런스 콜에서 생성형 AI 접목 계획에 대한 질문에 위의 이유로 보수적인 입장을 견지한 바 있다.

이러한 문제를 해결하기 위해 최근 논의되고 있는 것은 설명 가능한 AI (XAI: eXplainable Artificial Intelligence)다. 이는 생성형 AI가 도출한 결과물에 대한 이유를 사람이 이해할 수 있는 방식으로 제시하는 인공지능을 의미한다.

Samsung Securities

## 부록 2-3: 인터넷 어플리케이션 및 게임 - 서비스 개선, 운영 최적화, 그리고 비용 절감

## Potential &amp; Benefit

2023년 Chat GPT 출시 이후 기업과 정부 기관의 AI 도입이 가시화되고 있다. IDG에 따르면 기업들은 주로 AI를 개발, 보안 등 IT 업무나 고객센터 서비스에 가장 많이 적용할 계획이다. 기업들은 AI 적용으로 1) 기존 제품과 서비스 개선(55.3%), 2) 기업 내외부 운영 최적화(44.7%), 3) 비용 절감(27.8%) 등을 기대하고 있다. 실제로 생성 AI를 도입한 기업들은 84% 긍정적인 효과를 보고 있다고 밝혔다. 생성 AI 도입 효과에 대한 긍정적인 평가는 후속 투자 확대에 이어질 전망이다.

정부 및 공공기관 역시 행정 편의성 개선을 위한 AI 도입에 적극적으로 나서고 있다. 정부 및 공공 기관은 데이터 보안이 중요한 만큼 로컬 기업과의 협력으로 내부 AI 구축을 선호한다. 국내에서도 한국은행과 금융권이 네이버와 MOU를 체결하고 하이퍼클로바X를 업무에 적용할 계획이라 밝혔다.

해외 정부들도 AI 중심의 디지털 전환을 추진하고 있다. 각국 정부들은 디지털 주권과 보안 유지 차원에서 소버린 AI를 강조하고 있다. 이에 자국 AI 기술 확보에 투자하거나 안보적으로 위협이 되지 않는 국가의 기업에게서 AI 기술을 도입하고 있다.

**영상 콘텐츠 시장 본격화.** 오픈AI에서 텍스트를 영상으로 만들어주는 소라(SORA) 모델 공개로 AI의 영역이 이미지를 넘어 동영상으로 확대되었다. 소라의 등장으로 전문가들은 미디어, 엔터테인먼트, 게임 등 콘텐츠 산업 자체의 지각 변동을 예상하고 있다. 앞으로는 영화나 드라마, CF 등 영상물 제작에 대규모의 촬영 인프라가 필요 없어질 것으로 예상되기 때문이다. 영상 제작의 문턱이 크게 낮아지는 만큼 1인 미디어와 전문 프로덕션의 경계도 사라질 것으로 전망한다.

유튜브나 틱톡 등 영상 플랫폼 시장에도 변화가 생길 전망이다. 영상 콘텐츠 제작과 편집의 문턱이 사라짐에 따라 영상 콘텐츠의 양이 폭발적으로 늘어날 것으로 예상되며, 이 변화의 틈을 파고든 새로운 플랫폼의 출현도 충분히 가능해 보인다.

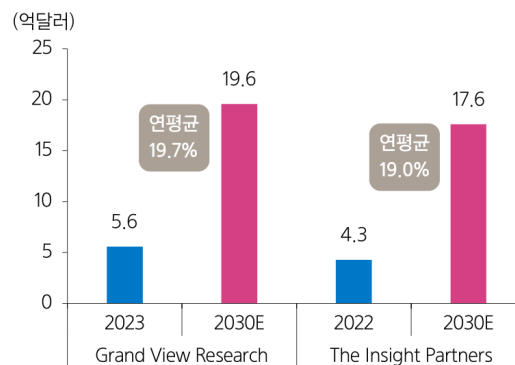
게임 시장에서도 변화가 불가피할 것으로 예상된다. 실시간 영상 제작이 가능해지며 게임과 영상의 혼합된 인터랙티브 무비 형태의 게임 제작이 늘어날 전망이다. 또한 게임 엔진 없이도 영상 내 객체의 제작과 조정이 가능해짐에 따라 게임 개발 비용이 크게 낮아질 것으로 예상되며, 이는 1인 제작사의 확장으로 이어질 것이다.

## 소라로 만든 3인칭 자동차 운행 영상



자료: 오픈AI, 삼성증권

## 영상 생성 AI 시장 전망



자료: Grand View Research, The Insight Partners, 삼성증권



**B2C 유료 구독 서비스 확대.** 지난 20년간 구글과 메타는 일반 소비자에게 무료로 서비스를 제공하는 대신 광고로 수익을 창출하는 모델을 고수해왔다. 그러나 증가하는 AI 추론 비용으로 플랫폼 업체들도 이용자 대상 직접 과금 모델 도입을 확대할 전망이다. 4월 4일 Financial Times는 구글이 기존 유료 구독 서비스에 AI 기반 검색을 추가하는 방안을 검토 중이라고 밝혔다.

이는 오픈AI의 챗GPT가 사용자의 질문에 완성된 답변을 내놓으며 이용자 만족도를 높이고 있는 상황과 무관치 않다. 오픈AI는 GPT의 구독 매출이 빠르게 증가함에 따라 23년 매출이 20억 달러를 돌파했다. 구글이 현재의 검색 시장 내 점유율을 유지하기 위해서는 생성 AI의 적용이 필요한 상황이고, 늘어나는 인프라 비용을 감당하기 위해서는 유료화가 불가피했을 것으로 추정된다.

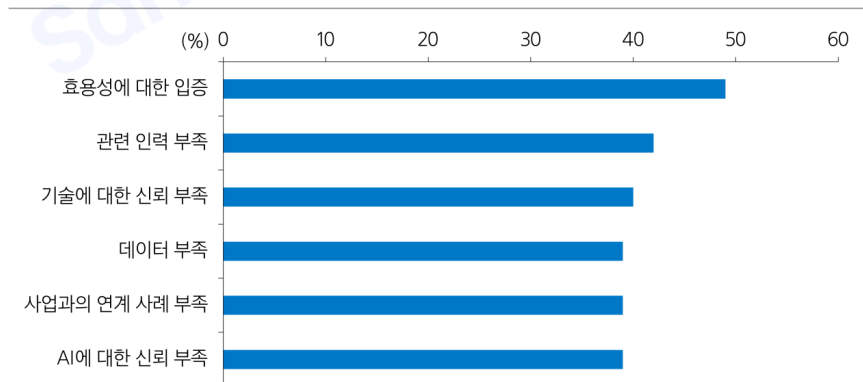
다행인 점은 AI 모델의 완성도 향상에 따라 일반 소비자들도 기꺼이 지갑을 열고 있다는 점이다. AI 챗봇에 멀티 모달 적용으로 텍스트뿐 아니라 이미지와 영상까지 검색과 생성이 가능해짐에 따라 B2C AI 서비스들의 구독료 또한 추가 상승이 가능할 것으로 전망한다.

#### Barriers to overcome

##### 생성 AI 기업 도입의 장애물

많은 기업들이 생성 업무에 생성 AI 도입을 검토하고 있으나, 여러 이유로 실제 도입으로 이어지는 비율은 48%에 불과하다(Gartner 설문조사, 2023). 이 중 가장 큰 장애물이 되는 건 생성 AI 도입의 효용성에 대한 입증이다. 생성 AI 기술을 기업에 도입하는데 한 두 분이 드는 게 아니고, 민감한 내부 정보 유출에 대한 우려도 있는 상황에서 생성 AI를 도입 효과를 정확히 예측할 수 없다는 점은 AI 도입을 주저하게 만든다. 또한 IT 회사가 아니라면 기업 내 데이터를 분류, 정제하고, 생성 AI를 적용시키는 과정 자체도 도전이 될 수 있다. 프로젝트 전반을 주도할 수 있는 기술 인력이 부족하다는 점도 기업들의 AI 도입의 장애물이 되고 있다. 이런 점을 고려하면 기업들의 AI 도입은 기업별 AI 커스터마이징을 통한 사용보다는 MS 코파일럿 365나 Adobe의 파이어플라이처럼 기존 어플리케이션에 내장된 생성 AI를 사용하는 형태로 우선 나타날 것으로 예상된다.

##### 기업 내 AI 도입을 가로막는 장애물



자료: Gartner (2023), 삼성증권

## 부록 2-4: 디바이스 - 기계와의 소통, 키보드에서 손가락으로, 손가락에서 새로운 Interface로

## Potential &amp; Benefit

## AI를 위해 디바이스에서 필수적인 기능

AI 기능을 활용하기 위해 적어도 지금 스마트폰보다 훨씬 많은 센서를 달고, 특히 생체 정보를 더 얻고, 내가 보고 듣는 것을 똑같이 보고 듣는 위치에 장착할 수 있는 디바이스가 필요하다. 디스플레이 역시 지금과는 다른 방식이 필요한데, 예를 들면 XR(Extended Reality)이 구현되는 형태가 선호될 것이다.

현재는 그림, 글의 생성이나 검색 기능에 국한되어 있지만, 앞으로는 내 주변 상황과 내 생체데이터의 확보가 필요하며, 디바이스의 발전은 내 주변 상황을 파악하는 도구로서 작동해야 한다.

## 스마트폰에 적용될 AI 기능이, 다음 디바이스의 가장 중요한 힌트



- 기계와 대화한다. (필요한 말은 물론 불필요한 말까지도 나눈다)
- 내가 처한 상황과 주변 상황을 파악한다.
- 내 건강과 스케줄을 챙긴다
- 내 사진과 문서를 검색해 준다.
- 나의 선호도와 관심사를 파악한다.
- 쇼핑을 도와 주고, 길을 찾아 준다.
- 그림을 그려 주거나 글을 대신 써 준다.
- 필요하면 검색도 해준다.

자료: 삼성증권

## 예상되는 킬러 서비스 (Application)

## (1) AI agent

AI agent는 마치 비서처럼 사용자에게 닥친 문제를 AI가 해결해준다는 개념이다. 기계와 내 일정과 계획을 조율하고 대화를 통해 감정을 기록하며 가벼운 대화를 나눌 수 있다. 기계가 상황을 먼저 파악해 조언을 받을 수도 있다.

이를 위해 기계는 대화를 자유롭게 나눌 수 있어야 하고, 해결책을 위해 해결 과제를 몇 가지의 Task로 쪼갤 수 있어야 한다. 마지막으로 해결책을 찾기 위해 무엇을 해야 하는지를 구할 수 있어야 한다. 언어 모델을 바탕으로 가장 합리적인 답을 추천하거나, 아니면 문제를 더 잘 해결할 수 있는 다른 머신러닝이나 다른 앱, 다른 AI를 부를 수도 있어야 한다. 이렇게 문제를 해결하는 과정에서 private data를 확보한다면, 더욱 완성도 높은 과제 수행을 할 수 있다.

## (2) AI 미디어의 출현

제일 먼저 상상할 수 있는 것은 소셜 네트워크용 그림을 생성하는 것이다. 메시지를 바탕으로 인스타그램에 업로드할 만한 그림을 생성할 수 있다. 더 나아가서는 AI가 만든 영화, AI가 만든 게임을 즐길 수 있다. 특정 셀러브리티 AI와의 대화를 즐길 수도 있다.

## (3) 지도와 쇼핑의 고도화

검색 엔진과 인터넷이 지도와 쇼핑을 바꾸었듯이, AI는 좋은 인프라 환경이 될 것이다. 우리가 표현하기 힘든 요구 사항들 - 예를 들면, 조용하며 개별 방이 있고 반경 200m의 음식점을 찾을 수 있다. 예전보다 훨씬 더 정교하게 검색할 수 있고, 검색과 지도, 쇼핑을 끊임 없이 이어 서비스를 만들 수 있다.

**Barriers to overcome****하드웨어의 기술 병목**

(1) **저전력 프로세서와 메모리:** CoWoS와 HBM처럼, AI의 성능 개선과 저전력을 동시에 달성할 수 있는 모바일 프로세서의 혁신이 아직 나타나지 않았다. 새로운 시대의 ARM core가 필요하다.

NPU의 도입과 사용은 전력 소모 문제를 야기한다. NPU가 달렸지만 잘 사용하지 않는다면 문제가 되지 않지만, 성공적으로 AI 서비스가 정착되어서 하루 종일 NPU를 써야 하는 상황이 온다면 배터리 소모가 클 수 있다. NPU뿐 아니라 GPU, DSP, ISP, CPU, 메모리까지 각종 반도체의 로드와 전력 소모가 커질 것이다. 비전 프로를 무선 기기로 만들지 못한 애플의 고민을 떠올려야 할 때다. 당장은 AI 연산에 필요한 배터리 소모량이 높아짐에 따라 x86과 ARM 사이의 성능 격차가 극대화되고 ARM으로 헤게모니가 넘어오는 결정적 계기가 될 것이라 생각한다. 하지만 저전력에 특화된 ARM조차도 AI 시대를 감당하기에 충분한 저전력 반도체는 아니다. 2002년 인텔 센트리노의 저전력 CPU가 8시간동안 무선으로 작동하는 노트북 PC를 만들었고, 이것이 노트북 PC 교체 주기의 원동력이 되었음을 상기해야 한다. 훨씬 더 저전력에서 돌아가는 AI 반도체가 등장하지 않으면, 주머니에 넣고 하루 종일 충전 없이 쓸 수 있는 AI 단말기를 만드는 것이 힘들어질 수 있다.

로컬에서 수행하는 계산이 어디까지인지 설정하는 것도 하나의 우회적인 방법이 될 수 있다. 로컬에서 작동해야 할 AI는 크게 두 가지일 것이다. 첫째, 대규모 언어 모델(LLM)의 일부를 잘라 경량화 모델로 변환하여 API로 로컬에 공급되고 나면 우리는 경량화 모델을 디바이스에서 운용하게 될 것이다. 추론용 반도체 사용이 빈번하게 나타날 것이라 예상된다. 경량 모델의 특성상 FP16, FP32(부동소수점)와 같은 스펙보다는 INT4, INT8(정수형)과 같은 스펙이 중요할 수 있다. 세부적인 숫자 표현만큼 메모리 대역폭 감소를 통한 연산 비용을 축소가 중요하기 때문이다. 둘째, 우리가 수집한 영상과 음성 정보들, 신체데이터 중 대부분은 쓸모가 없다. 이중 쓸모 있는 것을 고르고, 데이터 중 노이즈를 제거하여 당장 학습하기에 좋은 재료로 데이터를 가공하는 것을 디바이스에서 수행하게 될 것이다. 문제는 세부적으로 어디까지 잘라서 API 제공을 해야 할지, 어떤 로우 데이터를 센싱해야 할지 시행착오가 필요하다는 것이다. 그 전까지는 비효율적인 데이터와 반도체 사용을 감내해야 한다.

(2) **웨어러블 카메라와 센서들:** 주변의 상황을 카메라 포함 몇개의 센서를 어디에 달아야 할까? AI 서비스에 필요한 주변 정보를 파악하는데 최적화된 센서의 위치와 양을 모른다.

(3) **디스플레이:** AI 시대에 우리에게 정보를 어떻게 전달할 것인가? 가상 그래픽을 띄우는 Mixed reality는 구현해야 할 것인가도 중요하다.

(4) **배터리:** 전체가 용량을 해결할 수 있을까도 풀기 어려운 병목이다.

**소프트웨어의 기술 병목**

(1) **SLM (Small language model):** 더욱 더 작은 모델이 필요하다. 더 작으면서 LLM과 비슷한 성능을 내야 한다. SLM 기술이 없는 AI 서비스는 마치 MPEG 기술이 나오지 않은 비디오를 스트리밍하겠다는 이야기와 같다.

(2) **VLM (Vision language model):** 비전 기술이 필요하다. 주변의 상황과 주어진 Task를 이해하여 행동을 결정해야 한다. 처음에는 주인에게 조언하는 방식에서, 나중에는 로봇의 움직임을 결정할 것이다.

(3) **SSM (State space model):** 현재 상태를 완벽하게 파악하는 시가 필요하다. 예를 들면 자동차에 악셀을 밟았을 때 실제 속도계와 어떤 차이가 있는지 (측정값의 지연은 있는지), 악셀의 압력과 속도가 어떻게 변화하는지 실질 데이터를 계산할 수 있어야 한다.

## 부록 2-5: 헬스케어 - 궁극적인 인공지능의 게임 체인지, 건강하게 오래 살기

## Potential &amp; Benefit

## 인공지능이 탐하는 영역은 생명 활동에 대한 연구

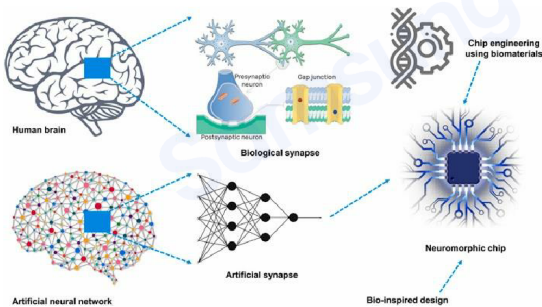
인간의 뇌를 넘어 AI가 탐하는 궁극적인 영역은 사람의 신체 내에서 일어나는 생명 활동에 대한 연구 일 것이다. AI의 발전으로 억만 장자들의 불로불사의 꿈은 전설 속의 이야기가 아니라 현실로 다가오고 있다. 레이 커즈와일이 꿈꾸는 2045년 미래는 뇌를 클라우드와 연결해 기억을 분산 저장하여 디지털 불로장생을 누리는 것이다. AI를 통해 인간의 뇌와 기계 신체가 합쳐진 미래가 그려지고 있다.

## 시냅스와 유사한 파라미터를 늘려 인간의 뇌에 가까워지는 인공지능

AI는 텍스트, 영상 분야 등에서 데이터를 학습하면서 인간의 뇌에 가까워지는 진화 과정에 있고, 2045년에는 인간의 지능을 넘어설 것으로 전망하고 있다. AI의 인공 신경망의 Parameter는 인간 뇌의 신경 세포 간의 상호 작용 통로인 시냅스와 유사하다. 인간의 뇌는 신경 세포가 1,000억 개가 있으며, 각각의 신경 세포는 다른 신경세포와 상호작용을 하면서 1,000조 개의 시냅스를 구축하여 기억, 학습, 인지 기능을 수행한다. AI는 Parameter를 늘림으로써 인간의 뇌처럼 생각하도록 발전하고 있다.

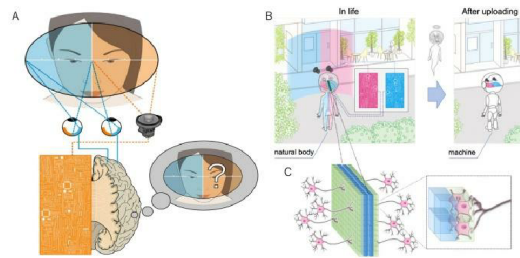
인간의 관찰과 유사하게 구사하는 로봇에 대한 연구도 진행되고 있다. 최근 발표된 로봇은 인공 지능을 통해 스스로 사고하면서 관절을 움직여 행동하는 모습을 보여줬다. 해당 로봇이 보여준 추론 과정과 판단에 따른 행동 양상은 인간의 뇌와 몸이 작동하는 과정을 그대로 모방하고 있다.

## 인간의 신경 시스템과 AI 비교



자료: Ahmed, 2023, 삼성증권

## 뇌를 클라우드-로봇과 연결



자료: Watanabe, 2023, 삼성증권

**현실적인 인공지능과 헬스케어, 건강하게 살기**

인간 뇌와 기계 몸이 합체된 미래도 좋지만 좀 더 현실적인 접근도 필요하다. 사람의 신체를 지금보다 오래 쓸 수 있도록 하는 방법에 대한 연구 (단순히 말하면 신약 개발)가 이루어지고 있다. 인간은 늙게 마련이고, 유전자 돌연변이나 신체 대사 문제로 암, 알츠하이머, 당뇨 등의 질병에 걸리게 된다. 질병에서 자유로워질 수 있다면 고통에서 벗어날 수 있을 것이며, 노화를 방지할 수 있다면 다시 젊어질 수 있을 것이다. 많은 연구자들은 시가 인간이 질병의 고통에서 빠르게 벗어날 수 있도록 도와줄 것으로 믿고 있다.

사람을 구성하는 세포의 수는 평균적으로 32조 개(남성 36조 개, 여성 28조 개)이며, 세포를 구성하는 주요 요소로는 단백질이 있다. 우리 몸에는 약 2만 개의 단백질이 각각 다른 기능을 수행하고 있으며, 그중 3-4천 개의 단백질이 돌연변이, 결핍, 과발현 등으로 질병을 유발하는 것이 알려져 있다. 현재까지는 기술적인 한계로 약 600개의 단백질에 대해 치료제가 개발되었다.

질병을 유발하는 모든 단백질에 대한 의약품들을 개발할 수 있다면 사람은 질병에서 자유로워질 수 있지 않을까? 문제가 있는 세포나 조직을 건강한 세포로 교체를 하면 훨씬 오래 살 수 있지 않을까? 질병과 노화에서 벗어나고 싶어하는 인간의 궁극적인 욕망을 해결하는 것이 시가 나아가야 하는 방향으로 생각되며, 시를 활용한 신약 개발은 앞으로 매우 빠른 속도로 이루어질 것으로 판단한다.

**단백질 구조 분석: AlphaFold의 혁신**

수작업을 통해 확인했던 단백질 구조를 예측했던 과거에서, 컴퓨터 모델링을 통해 예측할 수 있는 시대로 넘어왔다. 물론 20년 전부터 컴퓨터 모델링을 통한 단백질 구조 예측은 시도되었지만 정확도가 떨어졌다. 하지만 판도는 2018년 구글이 지원하는 AI 연구실 DeepMind에서 공개한 AlphaFold 덕분에 바뀌었다. AlphaFold는 모든 단백질에 대해 원자 단위로 예측된 구조를 제시하고 있으며, 다른 분자간의 상호작용까지 3D 구조로 표현해 준다. AlphaFold의 출시로 AI 신약 개발이 날개를 달았다. 질병의 주요 원인 단백질의 구조를 AlphaFold를 통해 보다 정확히 예측할 수 있게 되었다.

AlphaFold 개발 이전에는 3차원 단백질 구조를 예측하는 것은 어려운 과제였다. 주로 X-ray 결정학, 핵자기공명 분광학, 초저온-전자현미경을 이용하여 단백질 구조를 분석했는데, 이러한 방법으로 단백질 구조를 밝히려면 수개월에서 수년의 시간이 소요되었으며, 많은 비용이 들었다. 보다 정확한 단백질 구조를 빠르게 예측하기 위해 AI 기반 컴퓨터 시뮬레이션과 모델링을 통한 연구가 꾸준히 진행되었으며, AlphaFold가 단백질 구조 예측을 획기적으로 개선시킴으로써 신약 개발 속도도 가속화될 것으로 기대하고 있다.

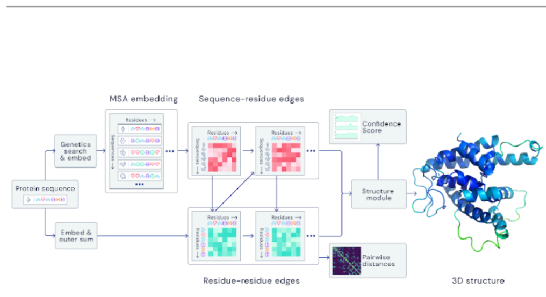
**합성 의약품 라이브러리**

이제 타겟 단백질과 결합할 수 있는 의약품의 개발이 필요하다. 고리화합물(Heterocycles) 구조는 FDA 승인을 받은 약물 및 현재 개발되고 있는 주요 합성 의약품에 포함되어 있다. 고리화합물은 고리 형태로 결합한 탄소 원자 사이에 비탄소 원자가 끼어 있는 구조로 추가적인 작용기를 결합해 다양한 구조의 의약품을 합성해 만들 수 있다. 고리의 모양이나, 작용기의 종류에 따라 다양한 단백질과 결합이 가능하다. 우크라이나 Enamine에서는 고리화합물 기반의 화학 라이브러리에 360억 개의 새로운 분자를 포함하는 Enamine REAL space를 보유하고 있다.

### 의약품 후보 물질 도출: 예측된 단백질 구조와 합성 의약품의 상호작용 확인

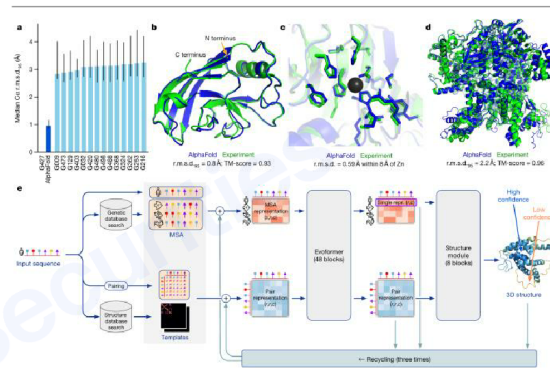
제약/바이오 기업은 Enamine Real에 화학 라이브러리를 구매한 후 AlphaFold로 구조 예측이 가능한 단백질과의 결합 여부를 시를 통해 확인한다. 화학 의약품과 타겟 단백질간의 상호 작용을 분석하여 매우 빠른 속도로 의약품 후보 물질을 도출해낼 수 있다. 하지만 선별한 화합물은 데이터상으로는 성능이 높지만 실제로 시험해보면 성능이 낮은 문제가 있으며, 성능 확인을 위해 시험, 검증 단계에서 12-18개월이나 소요되기 때문에 시간적 낭비를 줄이기 위해서는 예측 정확도를 높여야 한다. 개발사는 화합물의 2D뿐만 아니라, 3D 구조 데이터 추가하거나, 화합물의 종류를 늘리면서 데이터를 학습한다. 최근에는 합성 의약품뿐만 아니라 펩타이드, 항체 등의 의약품에 대해서도 AI 기술이 적용되면서 다양한 종류의 신약 개발에도 활발히 사용되고 있다.

#### AlphaFold: 단백질 구조 예측



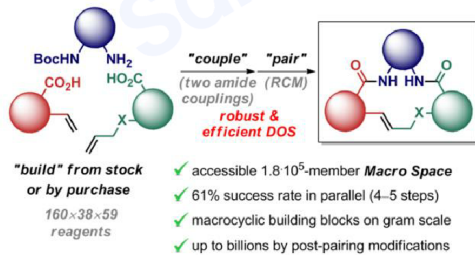
자료: DeepMind, 삼성증권

#### AlphaFold로 예측된 구조와 실제 단백질 구조 일치 여부 확인



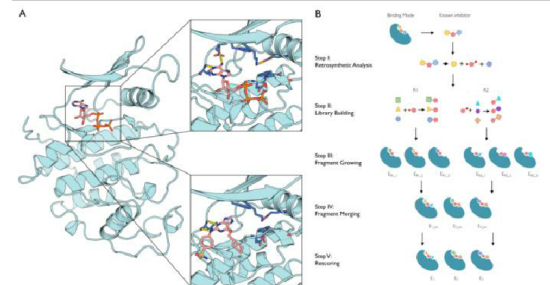
자료: Jumper et al., 2021, 삼성증권

#### 합성 의약품 라이브러리: 고리화합물 라이브러리 구축



자료: Enamine, 삼성증권

#### 단백질 Pocket에 결합할 수 있는 합성 의약품 스크리닝



자료: Andrianov et al., 2021, 삼성증권

#### AI를 통한 신약 개발 가속화

지난 60년 동안 미국에서 R&D 비용 10억 달러 당 승인된 약물의 수는 9년 마다 절반으로 줄었다. 현재 새로운 약을 시장에 출시하려면 10억 달러 이상의 자금과 10년 이상의 시간이 필요하다. 그러나 AI를 활용한 신약 개발은 박사 과정에서 5년간 수행한 실험을 15분으로 단축시켜줄 정도로 빠르다. 수제 신약 개발에서 자동화 과정으로 옮겨가고 있는 것이다. 타겟 스크리닝, 최적화 과정 등 신약 개발의 가장 앞 단계부터 AI를 활용하게 되면 최소 3-7년의 시간을 단축할 수 있다. 제약사들은 개발 단계부터 비용 절감을 위해 후보 물질 하나당 대략 2-4억 달러 규모로 파트너십을 체결하고 있다.



## Barriers to overcome

## 인공지능이 해결하기 어려운 임상

하지만 여전히 AI로는 해결될 수 없는 영역이 있다. 사람 대상으로 수행되는 임상 1-3상에서는 AI의 역할이 크지 않다. 약물의 투약 후 효과를 확인까지 절대적인 시간이 필요하기 때문에 임상 시험에서 AI를 통한 시간 단축은 다소 어렵다. 임상 시험에서 환자에서 약물의 효과를 확인하는 것 외에 AI를 활용하여 시간을 단축할 수 있는 부분이 있다.

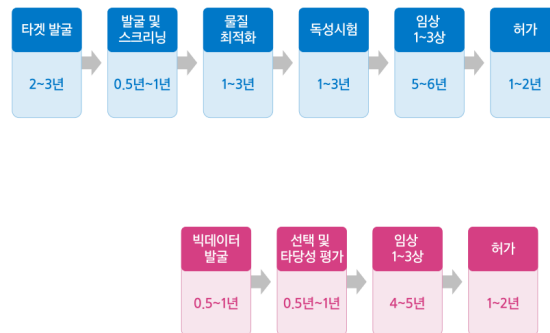
- **임상 디자인:** 어떤 환자에게, 얼마나 많은 환자에게 약을 투약할지, 어떤 데이터를 수집해야 할지 등에 대해 AI 분석이 이루어지고 있다. Jimeng Sun 연구실에서는 약물 구조, 질병, 임상에 참여한 환자 기준을 기반으로 임상 시험의 성공 여부를 예측하는 HINT 알고리즘을 개발했다. 예측된 결과에 따라 임상 디자인을 변경하거나, 개발을 중단할 수 있다.
- **임상에서 많은 시간이 소요되는 것은 환자 모집이다.** 임상 데이터를 활용하여 특정 질환에 대한 임상을 진행할 때 배제할 수 있는 환자의 기준을 찾아야 한다. James Zou의 Trail Pathfinder 시스템은 임상에 적합한 환자 기준을 완화하여 빠른 스크리닝을 통해 임상 속도를 높일 수 있다. 임상에 필요한 환자 수를 줄이거나, 환자가 원하는 임상 시험을 찾는 것도 AI를 통해 개선되고 있다.
- **임상 결과 분석:** Intelligent Medical Objects에서는 OpenAI의 GPT-4를 사용하여 임상 결과에서 안전성과 효능 정보를 추출하는 SEETrials를 개발했다. 임상을 디자인하는 연구자는 다른 연구자의 임상 디자인, 임상 결과 등을 빠르게 확인할 수 있다.
- **실제 임상 시험 성공률을 높이는 것이 중요하며, AI를 통해 발굴한 후보 물질의 PoC(Proof of Concept) 임상 결과에 주목해야 한다.** 2024-2025년 Schrodinger, Recursion 등에서 AI 기반으로 찾아낸 후보 물질에 대한 임상 결과 발표가 예정되어 있다. 긍정적인 임상 결과가 나올 경우, 글로벌 제약사들은 보다 적극적으로 AI 신약 개발에 참여할 것으로 예상된다.

## AI 신약 개발 업체와 글로벌 제약사와의 협업

시기	AI 신약 개발사	파트너사	규모 (USDm)	계약 내용
2023.07	Recursion	Nvidia	50	신약 후보 물질 발굴 AI 모델 개발
2023.05	XtalPi	Eli Lilly	250	비공개 타깃
2022.11	Insilico	Sanofi	1,222	6개 타깃
2022.1	Schrodinger	Eli Lilly	450+	비공개 타깃
2022.08	Atomwise	Sanofi	1,020	5개 타깃
2022.01	Generate	Amgen	1,900	최대 10개 타깃 단백질 치료제
2022.01	Exscientia	Sanofi	5,300	최대 15개 항암/면역질환 타깃
2021.12	Recursion	Roche	450	최대 40개 뇌질환/항암 타깃
2021.05	Exscientia	BMS	1,200	항암제, 자가면역질환
2020.12	Relay	Genentech	795	RLY-1971, GDC-6036 병용개발
2020.11	Schrodinger	BMS	2,755	2개 후보 물질 + 비공개 타깃
2020.09	Recursion	Bayer	1,030	폐/심장/신장 섬유화 등
2020.01	Exscientia	Bayer	267	심혈관계/항암제

자료: 삼성증권

## AI 신약 개발 시에 초기 연구 시간 단축



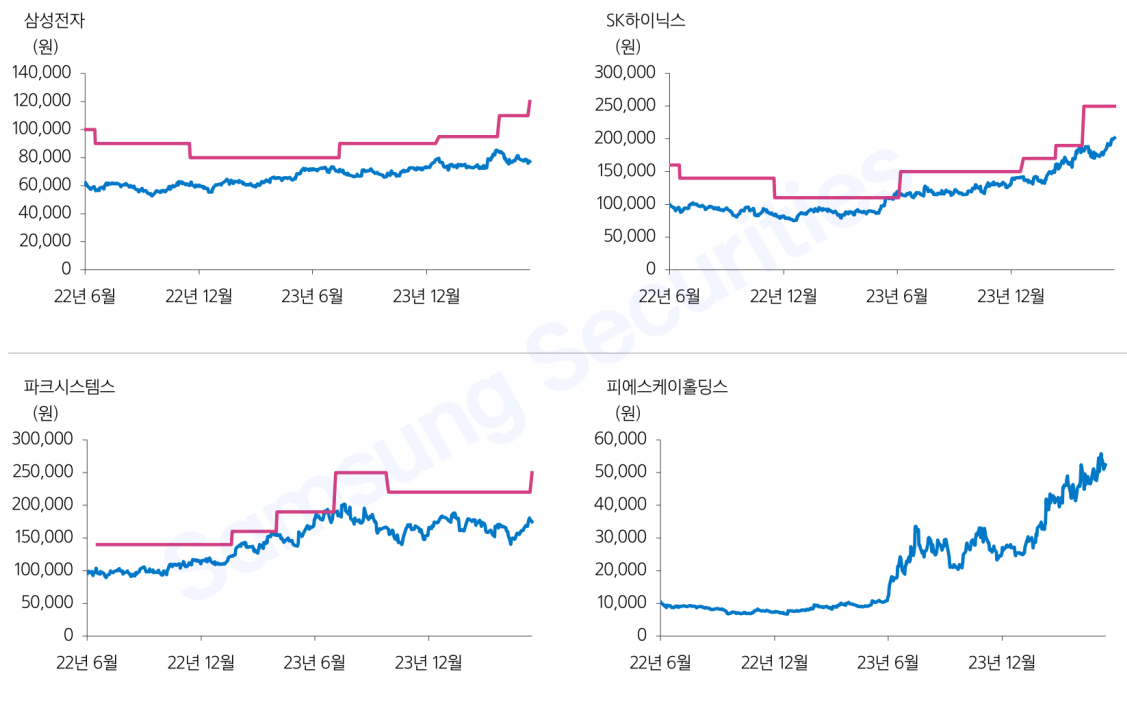
자료: 삼성증권



## Compliance notice

- 당사는 2024년 5월 27일 현재 삼성전자와(과) 계열사 관계에 있습니다.
- 본 조사분석자료의 애널리스트는 2024년 5월 27일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 2024년 5월 27일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자자 등 제3자에게 사전 제공된 사실이 없습니다.

## 2년간 목표주가 변경 추이



최근 2년간 투자의견 및 목표주가 변경 (수정주가 기준)

삼성전자							
일 자	2022/5/31	6/30	11/29	2023/7/27	2024/1/2	4/8	5/28
투자의견	BUY	BUY	BUY	BUY	BUY	BUY	BUY
TP (원)	100000	90000	80000	90000	95000	110000	120000
과리율 (평균)	-38.66	-34.78	-19.23	-22.05	-20.74	-28.36	
과리율 (최대or최소)	-33.20	-30.11	-8.25	-12.78	-10.21	-23.55	
SK하이닉스							
일 자	2022/6/15	6/30	11/29	2023/6/19	2024/1/2	2/23	4/8
투자의견	BUY	BUY	BUY	BUY	BUY	BUY	BUY
TP (원)	160000	140000	110000	150000	170000	190000	250000
과리율 (평균)	-41.14	-34.45	-18.57	-18.20	-17.87	-10.86	
과리율 (최대or최소)	-39.13	-26.79	8.64	-5.67	-7.94	-1.05	
파크시스템스							
일 자	2022/6/30	2023/2/1	4/14	7/17	10/10	2024/5/28	
투자의견	BUY	BUY	BUY	BUY	BUY	BUY	
TP (원)	140000	160000	190000	250000	220000	240000	
과리율 (평균)	-25.25	-12.47	-13.45	-27.74	-24.87		
과리율 (최대or최소)	-12.50	-1.44	1.89	-19.20	-14.41		
피에스케이홀딩스							
일 자	2024/5/28						
투자의견	Not Rated						
TP (원)							
과리율 (평균)							
과리율 (최대or최소)							

투자기간 및 투자등급: 삼성증권은 기업 및 산업에 대한 투자등급을 아래와 같이 구분합니다.

\* 2023년 7월 27일부터 기업 투자 등급 기준 변경

기업

BUY (매수)    향후 12개월간 예상 절대수익률 15% 이상  
                  그리고 업종 내 상대매력도가 평균 대비 높은 수준  
HOLD (중립)    향후 12개월간 예상 절대수익률 -15%~15% 내외  
SELL (매도)    향후 12개월간 예상 절대수익률 -15% 이하

산업

OVERWEIGHT(비중확대)    향후 12개월간 업종지수상승률이 시장수익률 대비 5% 이상 상승 예상  
NEUTRAL(중립)            향후 12개월간 업종지수상승률이 시장수익률과 유사한 수준 (±5%) 예상  
UNDERWEIGHT(비중축소)    향후 12개월간 업종지수상승률이 시장수익률 대비 5% 이상 하락 예상

최근 1년간 조사분석자료의 투자등급 비율    2024.03.31

매수(77.6%) 중립(22.4%) 매도(0%)

Samsung Securities

## 삼성증권

### 삼성증권주식회사

서울특별시 서초구 서초대로74길 11(삼성전자빌딩)  
Tel: 02 2020 8000 / [www.samsungpop.com](http://www.samsungpop.com)

삼성증권 Family Center: 1588 2323

고객 불편사항 접수: 080 911 0900



Member of  
**Dow Jones  
Sustainability Indices**  
Powered by the S&P Global CSA