AUDUN WICKSTRAND IVERSEN
Porteføljeforvalter

EMILIE KRUTNES ENGEN
Porteføljeforvalter

LEO RUNDGREN OLSEN
Junior Porteføljeforvalter

Hensikten med Disruptive Perspektiver:

Når vi analyserer ulike temaer bruker vi mye tid og mange verktøy (kvartalsrapporter, analyser, dialog med selskapene, bedriftsbesøk, excel, kalkulator og ordmodeller). Ofte lager vi små notater, og noen ganger store notater som vi tenker på som perspektiver. Vi er gamle nok til å vite at det sjeldent finnes sannheter, ofte bare ulike perspektiver.

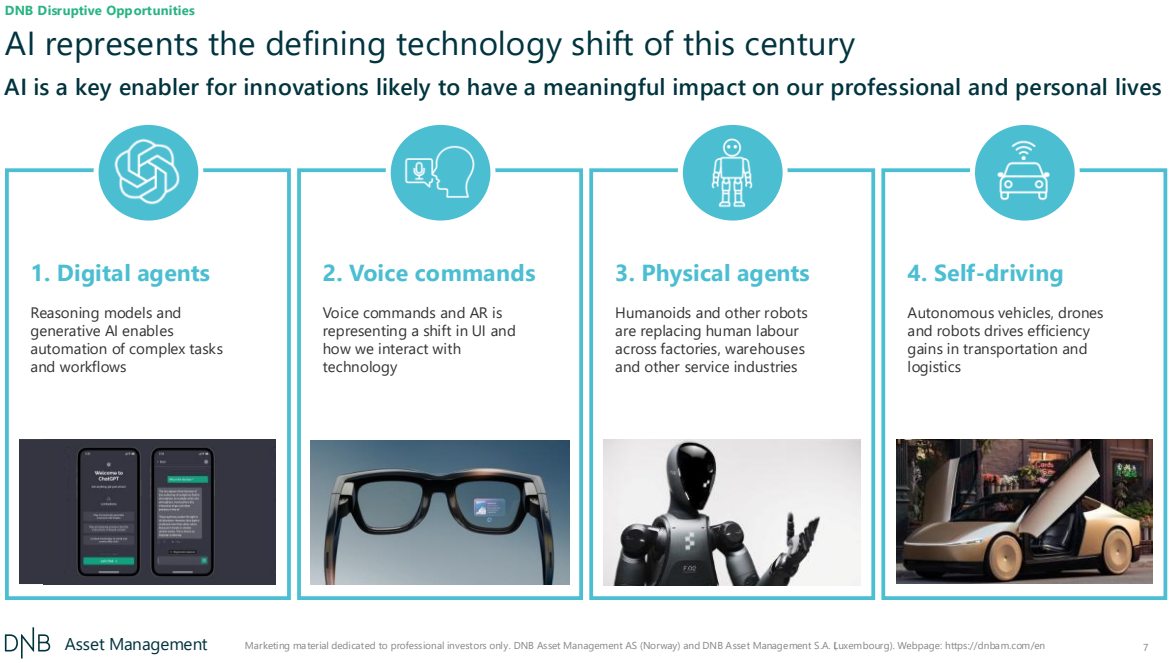
Disruptive Perspektiver har kun én hensikt: Å dele våre perspektiver på temaer som former vår fremtid. Dette er ikke akademiske notater, innlegg til et leksikon eller anbefalinger om å gjøre noe, kjøpe eller selge noe. Kun god gammeldags informasjonsdeling for å synliggjøre hvordan vi ser på ulike temaer på publiseringstidspunktet. Perspektiver blir ikke mindre, kanskje heller mer, når man deler det. Med det utgangspunktet; ha en fin reise i våre perspektiver.

Innholdsfortegnelse

Disclaimer

Fra biologisk hukommelse til symbiotisk minne

Hva trenger en humanoid for å bli virkelig nyttig? Eller hva må til før AI-brillene på nesen faktisk kan bli et ekstra lag av hukommelse? Svaret handler mye om minne.



Vi har de siste årene skrevet om de fire rytterne som fire ulike innovasjonsplattformer for å beskrive de største teknologiske skiftene i dette disruptive tiåret som vi er midt inne i. Digitale agenter begynner å handle på våre vegne i den digitale verden. Fysiske agenter er den fysiske forlengelsen av dette. Selvkjørende systemer beveger seg ut i den virkelige verden, mens stemmen gradvis blir et naturlig grensesnitt mellom menneske og maskin. Men jo mer autonome disse systemene blir, desto tydeligere oppstår en felles utfordring; forståelse av en og flere situasjoner over tid. De må huske, og ha kontinuerlig kontekst.

En humanoid kan ha verdens beste AI-modell, men den blir begrenset dersom den ikke klarer å holde sensorinformasjon og verdensmodellen aktiv mens den beveger seg gjennom landskapet. En digital agent blir lite personlig om den glemmer hva dere snakket om forrige uke. En selvkjørende bil må tolke enorme mengder

informasjon raskt nok til å reagere før verden rundt den endrer seg. Derfor er minne mer enn bare lagringen.

I datasenteret har HBM blitt en av de viktigste flaskehalsene for AI. På edge'n vokser behovet for minne som bruker mindre strøm og beholder informasjon uten konstant tilgang til skyen. Samtidig forsøker nye arkitekturer å flytte selve beregningen nærmere dataene. Minne går dermed fra å være en passiv komponent til å bli en stadig viktigere del av hvordan intelligensen faktisk fungerer.

Finnes det en parallell til oss mennesker?

Menneskelig hukommelse er verken perfekt eller statisk. Vi rekonstruerer fortiden. Følelser påvirker hva som setter seg. Erfaring blir over tid komprimert til noe vi kaller intuisjon. Bevisstheten velger bare ut en liten del av alt dette i øyeblikket.

Maskinene fungerer annerledes. De kan lagre med enorm presisjon og hente informasjon raskere enn vi noen gang vil klare. Men hittil har de vært overraskende dårlige til å bygge den typen kontinuerlig hukommelse som gjør et menneske til det samme mennesket fra én dag til den neste.

Nå begynner dette derimot å endre seg. Når agenter får varig, *non-volatile* minne (kommer tilbake til dette i kapittel 1.2), briller kan huske deler av livet vårt, og roboter bygger erfaring fra verden de faktisk opererer i, oppstår et nytt lag mellom biologisk og teknologisk hukommelse. Vi kan kalle det **symbiotisk minne**.

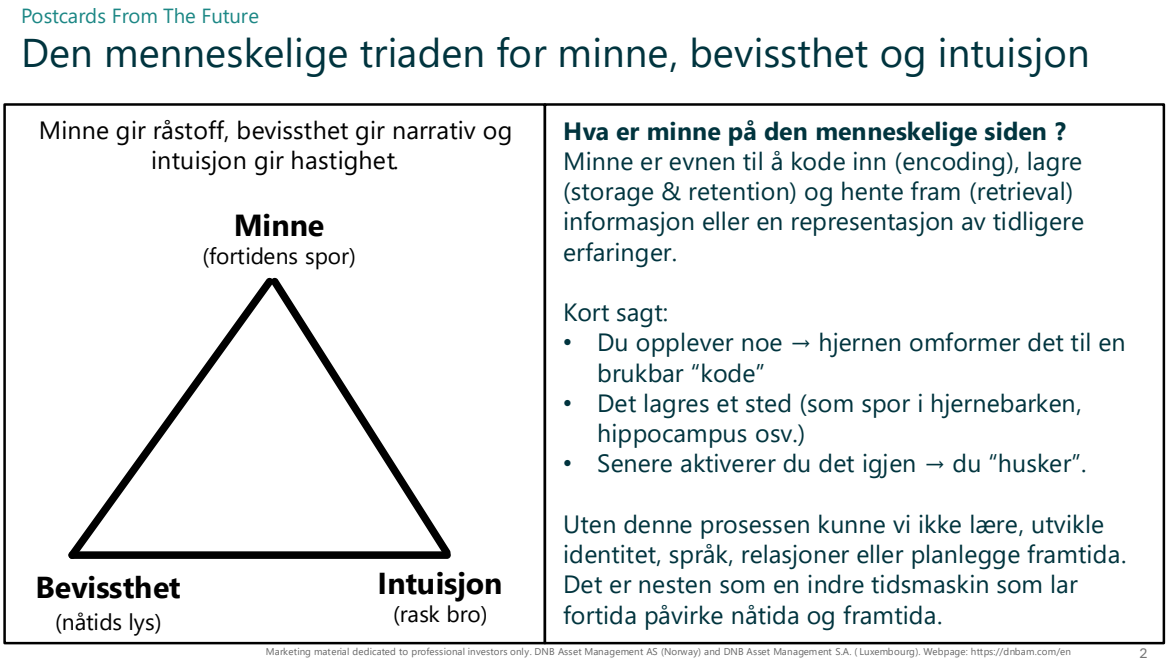
Det er i dette skjæringspunktet perspektivnotatet begynner. Først i vår egen hjerne. Deretter ned i silisium, HBM, NAND og nye minneteknologier. Videre til in-memory computing, edge AI og de fire rytterne. Til slutt lander vi i kanskje det viktigste spørsmålet for oss aksjefolk akkurat nå:

Har minneindustrien bare gått inn i enda en klassisk boom-bust-syklus, eller er AI i ferd med å gjøre minne til strategisk infrastruktur i en langt større teknologisk supersyklus?

For å forstå hvor vi skal, kan vi derfor starte med det eldste minnesystemet vi kjenner, nemlig vårt eget.

1.0 Den menneskelige triaden

En kjapp måte å tilnærme seg den menneskelige triaden er vist i figuren under.



Uten denne prosessen kunne vi ikke hverken lære, utvikle identitet, språk, relasjoner eller planlegge framtiden.

1.1 Biologisk minne

Minne ligger ikke på ett bestemt sted i hjernen, som en fil på en harddisk. Det likner mer på et nettverk. Ulike deler av hjernen bidrar med ulike funksjoner, og sammen skaper de det vi opplever som hukommelse.

Hippocampus er særlig viktig når vi danner nye minner. Den hjelper oss å lagre fakta, hendelser og opplevelser. Det er derfor skader i dette området ofte gjør det vanskelig å huske nye ting, selv om gamle minner fortsatt kan ligge der. Ved Alzheimers ser man ofte at hippocampus rammes tidlig. Da forsvinner gjerne de nyere minnene først, mens sterke minner fra barndom og ungdom kan henge igjen lenge.

Amygdala gir minnene følelsesmessig farge. Den er grunnen til at noen øyeblikk setter seg ekstra hardt. Et traume. En forelskelse. En smell. En lukt fra et gammelt hus. Tenk på det litt som at følelser fungerer som en markeringspenn i hjernen. Det som betyr noe markeres.

Prefrontal cortex, eller frontallappen på godt norsk, hjelper oss med arbeidsminne. Det er det vi holder aktivt i hodet akkurat nå. Et telefonnummer før vi skriver det ned. En setning vi forsøker å formulere. Flere ting vi prøver å balansere samtidig. Dette minne er kortvarig og sårbart. Blir vi avbrutt, kan det forsvinne på sekunder. Alle som har gått inn i et rom og glemt hvorfor de gikk dit, har kjent arbeidsminnes ydmykende kraft. Derfor forsøker vi alltid ha en de-brief etter møter/selskapsbesøk

Så har vi cerebellum (lillehjernen) og basalgangliene, som er viktige for prosedyreminne, balanse og situasjonsforståelse. Dette er det vi ofte kaller muskelminne, selv om minne egentlig ikke sitter i musklene. Det sitter i nervesystemets mønstre. Derfor kan du sykle, skrive på tastatur, knyte sko eller kjøre bil uten å tenke gjennom hver enkelt bevegelse. Kroppen husker det.

Biologisk sett handler minne ofte om hvor i denne kjeden problemet oppstår. Noen ganger klarer vi ikke å lagre nye minner. Andre ganger ligger minne der, men vi får det ikke frem. Og noen ganger er selve lagringen skadet. Det er forskjellen på å for eksempel aldri skrive ned notatet eller å miste nøkkelen til arkivet.

Vi kan også dele hukommelsen inn etter hvor lenge den varer og hva den brukes til. Sensorisk minne varer bare i et øyeblikk. Det er ekkoet av det du akkurat hørte, eller bildet som henger igjen et lite sekund etter at du lukker øynene. Arbeidsminne varer

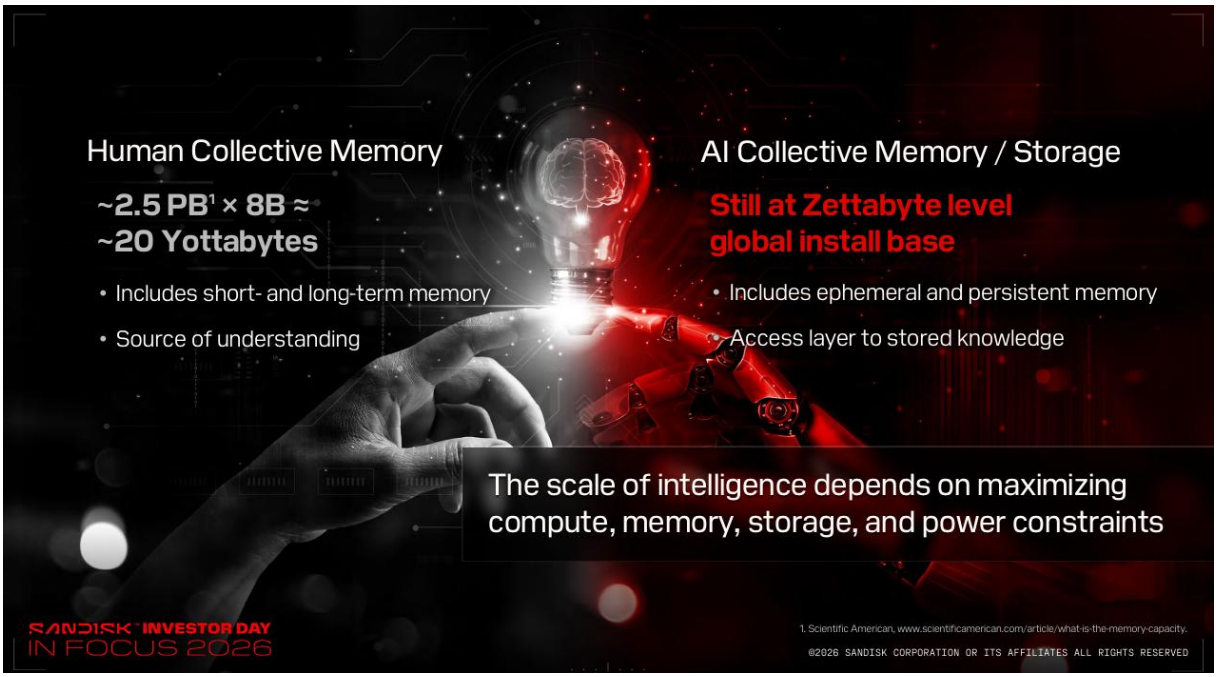
litt lenger, ofte bare 20-30 sekunder uten repetisjon. Det er begrenset, og nyere forskning tyder på at vi ofte klarer færre enheter enn den gamle tommelfingerregelen om 7 (+/- 2), særlig hvis informasjonen ikke gir mening for oss.

Langtidsminne er større og rikere. Der finner vi det deklarative minne, altså det vi kan sette ord på. Det deles gjerne i episodisk minne og semantisk minne. Episodisk minne er personlige hendelser, som bursdagen din i fjor eller første skoledag. Semantisk minne er fakta, som at Oslo er hovedstaden i Norge. I tillegg har vi det implisitte minne, som handler om ferdigheter og vaner vi kan utføre uten å tenke bevisst på dem.

Noe som også er veldig interessant er det at hukommelse ikke bare handler om lagring. Det handler også om identitet. Vi er, på mange måter, en fortelling hjernen stadig skriver om oss selv. Hva vi husker, hva vi glemmer, og hvilke minner følelsene våre forsterker, former hvordan vi forstår hvem vi er. Minne er derfor ikke kun et biologisk arkiv. Det er et levende system som binder sammen fortid, nåtid og forventningen om fremtid.

1.2 Teknologiske minne

Når vi flytter oss fra hjernen til dataverdenen, får ordet minne en litt annen betydning. Her handler det ikke om barndom, lukt, traumer eller identitet. Det handler om hvor raskt en maskin klarer å holde data tilgjengelig, hente den frem og skrive dem tilbake.



Kilde: Sandisk Investor Day 2026

Teknologisk minne er kapasiteten til å lagre data slik at de kan brukes av en prosessor, en AI-modell eller et datasystem. Noe minne er midlertidig og forsvinner når strømmen slås av. Det kalles *volatile memory*, som DRAM (Dynamic Random Access Memory) og HBM (High-Bandwith Memory). Dette er arbeidsbenken til maskinen. Raskt og dyrt, men tett på beregningen (compute) under inferens.

Annet minne er permanent og blir liggende selv om strømmen forsvinner. Det kalles *non-volatile memory*, som flash, SSD-er og harddisker. Dette er mer som arkivet. Den enkle måten å forstå det på er at jo raskere minne er, desto nærmere må det ligge selve beregningen. Og jo nærmere det ligger, desto dyrere og mer begrenset er det.

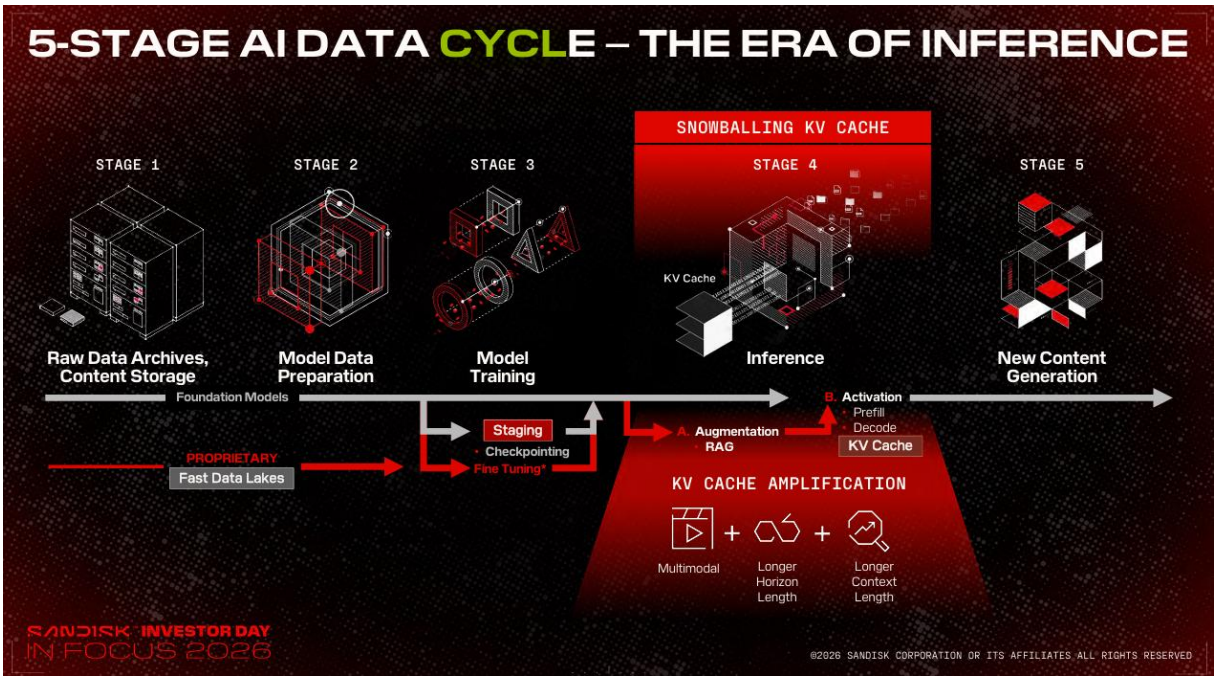
Minne type	Eksempel plassering	Varighet	Hastighet	Volatilitet	Analog til menneske?
Register / L1-cache	Inne i CPU	Nanosekunder	Ekstremt rask	Volatile	Umiddelbar sansning /arbeidsminne
SRAM cache (L2/L3)	CPU/GPU cache	Mikrosekunder	Veldig rask	Volatile	Korttidsminne/ aktivt fokus
DRAM (vanlig RAM)	Hovedminne i PC/server	Millisekunder	Rask	Volatile	Korttids- / mellomlangsiktig hukommelse
HBM (High Bandwidth Memory)	AI-GPU-er (NVIDIA H200, Blackwell, etc.)	Nanosekunder	Ekstremt rask	Volatile	Super-rask "arbeidshukomm else" for AI
NAND Flash / SSD Lagring	(NVMe, etc.)	År – tiår	Middels	Non-volatile	Langtidsminne (men tregere henting)
Optisk / tape	Arkiv	Tiår – århundrer	Svært treg	Non-volatile	"Gammelt arkivminne"

Derfor bygges datamaskiner som et hierarki. Innerst ligger cache, så HBM eller DRAM, deretter SSD-er og til slutt billigere langtidslagring. Det ligner litt på hvordan vi mennesker fungerer. Det du holder i hodet akkurat nå er lett tilgjengelig, men begrenset. Det du må grave frem fra langtidsminne tar mer tid.

Minne i denne AI-æraen har blitt en knapp ressurs. Modellene blir ikke bare bedre og/eller større, men de må også holde enorme mengder informasjon aktivt tilgjengelig mens de svarer, resonnerer og handler. Når en språkmodell genererer tekst, må den hente frem parametere, mellomregninger og kontekst. I store modeller oppstår derfor en kamp om minnebåndbredde, altså hvor raskt data kan flyttes inn og ut av beregningen samtidig. Derfor trenger man chips som får nok data raskt nok, fra minne.

Det er en pararell til menneskelig biologisk minne. I hjernen handler encoding om at vi danner et nytt minne. I dataverdenen betyr det at data skrives inn, for eksempel når en modell trenes og parameterne oppdateres. Lagring handler om at informasjon holdes tilgjengelig, enten i aktivt minne under inferens eller i treningssjekkpunkter på disk. Henting handler om å få informasjonen frem igjen. Hvis dette tar for lang tid, stopper hele flyten opp.

Det er også grunnen til at KV-cache har blitt så viktig i store språkmodeller. Modellen må holde styr på tidligere deler av samtalen raskt nok til å kunne svare videre uten å regne alt på nytt. Minne blir en del av selve ytelsen og ikke bare et sted man lagrer data.



Kilde: Sandisk Investor Day 2026

Men forskjellen mellom menneskelig og teknologisk minne er kanskje enda mer interessant enn likhetene. Menneskelig minne er levende. Vi assosierer det ofte med noe, følelser påvirker og det kan være litt upresist. Vi husker ikke verden som et kamera. Vi rekonstruerer den. Minnene våre endrer seg over tid, blandes med følelser og påvirkes av hvem vi er når vi henter dem frem. Det samme minne kan føles annerledes ti år senere.

Teknologisk minne er motsatt. Det er presist, statisk og bit-nøyaktig. En fil som lagres riktig, hentes frem likt hver gang. Det er styrken. Men også svakheten. Teknologisk minne forstår ikke hva det lagrer. Det vet ikke om en tekst er vakker eller om en opplysning betyr noe. Det kan reprodusere informasjon, men ikke gi den mening eller kontekst av seg selv.

Derfor prøver moderne AI-systemer å bygge bro mellom lagring og forståelse. RAG-systemer henter relevant informasjon fra eksterne databaser. *Fine-tuning* former modellen over tid. Agentminne gjør at digitale agenter kan huske preferanser, tidligere handlinger og kontekst. Likevel er det viktig å skille mellom å lagre og å forstå. En maskin kan hente frem alt. Det betyr ikke at den husker slik vi gjør.

Menneskelig minne er rotete, men meningsfullt. Teknologisk minne er presist, men tomt for mening før det kobles til kontekst og handling. Det er i dette skjæringspunktet mye av den neste AI-utviklingen skjer. Ikke bare i større modeller, men i bedre minne. Raskere minne. Billigere minne. Og etter hvert systemer som ikke bare svarer ut fra det de har lært, men også fra det de faktisk kan hente, lagre og bruke i øyeblikket.

1.3 Hvorfor nå?

Akkurat nå er høybåndbreddeminne, eller HBM, en av de store flaskehalsene i AI. Moderne AI-modeller trenger ikke bare regnekraft. De trenger enorme mengder raskt tilgjengelig minne. Når modellene skal håndtere stadig lengre kontekster, og hundretusener eller millioner av tokens, må de holde mye av samtalen og beregningen aktive samtidig. KV-cache alene kan i de mest krevende tilfellene spise enorme mengder minne under inferens.

Det er derfor HBM har blitt så viktig, fordi AI-modeller uten rask tilgang til data blir som en Formel 1-bil med sugerør inn til bensintanken. Motoren kan være brutal, men den får ikke nok drivstoff raskt nok. HBM løser noe av dette ved å plassere minne tett på prosessoren og gi ekstrem båndbredde. Ulempen er riktignok at det er dyrt, krevende å produsere og tar plass i produksjonskapasiteten. Når hyperscalerne fyller ordrebøkene med AI-behov, merker resten av verden det gjennom dyrere og knappere vanlig minne.

Mens menneskelig minne blir mer verdsatt fordi det er kreativt, narrativt, følelsesladet og uperfekt, har data og teknologisk minne blitt den nye oljen. Knapp og helt avgjørende for skaleringen av digitale agenter, fysiske agenter, selvkjørende systemer og tale som grensesnitt. Det er ikke nok at AI kan regne. Den må også huske, raskt nok og billig nok.

Samtidig er ikke minne det samme som bevissthet. Minne er evnen til å lagre, holde på og hente frem informasjon eller erfaringer fra fortiden. Det kan være bevisst, men trenger ikke være det. Du kan sykle uten å tenke på hvordan du holder balansen.

Bevissthet er noe annet. Det er opplevelsen av å være til stede akkurat nå. Den indre scenen. Følelsen av at noe faktisk skjer for deg. At rødt ser ut som rødt. At kaffe lukter kaffe. At smerte føles som smerte. Filosofene kaller ofte dette *qualia*, men vi kan si det enklere. Bevissthet er ikke bare informasjon. Det er opplevelsen av informasjon.

En enkel huskeregel er denne. Minne er fortidens erfaring. Bevissthet er nåtidens tilstedeværelse.

Du kan ha minne uten bevissthet. Vaner, reaksjoner, priming og ferdigheter skjer ofte i bakgrunnen. Du kan også ha bevissthet uten å hente frem et tydelig minne. Du kan sitte i solen, høre en lyd, kjenne vinden i ansiktet og bare være i øyeblikket, uten å koble det til en bestemt episode fra fortiden.

Intuisjon ligger et sted mellom disse to. Den føles ofte som en plutselig innsikt, men bygger som regel på lag av erfaring vi ikke klarer å forklare steg-for-steg. En investor kjenner at noe skurrer i et case før regnearket viser det. En fotballspiller vet hvor pasningen skal slås før han rekker å tenke. Intuisjon er nesten magi, men ikke helt. Det er komprimert erfaring som dukker opp som en følelse. Ja, vi kan kalle det mageføelsen.

Det gjør koblingen mellom minne, bevissthet og intuisjon så interessant. Minne gir oss materialet. Bevissthet lyser opp noe av det i øyeblikket. Intuisjon bruker

mønstrene under overflaten og sender opp et signal før den rasjonelle forklaringen er ferdig skrevet.

For mennesker er dette dypt biologisk og personlig. For maskiner er det foreløpig noe annet. AI-systemer kan lagre enorme mengder data, hente dem lynraskt og finne mønstre mennesker aldri ville sett. Men det betyr ikke at de opplever. De har teknologisk minne, og de kan etterligne intuisjon gjennom mønstergjenkjenning, men de har ikke nødvendigvis en bevissthet.

Kanskje er det nettopp her skillet går. Teknologisk minne handler om fart og presisjon. Menneskelig minne handler også om mening. Bevissthet handler om opplevelse. Intuisjon handler om dyp erfaring som ikke kan forklares.

Hvis HBM er den nye oljen for AI, er menneskelig minne fortsatt noe merkeligere og rikere. Det er lagret erfaring, men også historien vi forteller oss selv om hvem vi er.

Kort oppsummert

1. Minne gir innholdet og mønstrene.
2. Bevissthet gir muligheten til å oppleve, reflektere over og fortelle om dem.
3. Intuisjon er den lynraske, følelsesladede syntesen av minne som ofte skjer utenfor bevisstheten lys, men som likevel kan guide oss svært presist.

De tre danner en slags triade. Minne er drivstoffet, bevisstheten er føreren, og intuisjonen er den automatiske navigasjonen.

1.4 Triangelet og de fire rytterne

La oss koble dette tilbake til de fire agentene våre. Digitale agenter, fysiske agenter, selvkjørende agenter og relasjonsorientert tale.

Minne, bevissthet og intuisjon er tre av de mest grunnleggende byggesteinene i menneskelig tenkning. Digitale agenter har foreløpig noe ganske annet. De har kontekstvindu, cache, databaser og verktøy. De kan holde mye informasjon aktivt i en

samtale, men mangler ofte ekte kontinuitet over tid. Mange systemer starter fortsatt litt på nytt hver gang. De kan hente fra dokumenter, bruke RAG, lagre preferanser og bygge *persistent state*, men de har ikke et menneskelig episodisk minne.

Og grensen begynner å flytte seg. Når en digital agent får tilgang til tidligere samtaler, kalender, e-post, dokumenter, bilder, lokasjon og løpende kontekst, blir den mer enn en chatbot. Det blir en ekstern hukommelse. Den kan huske hva du sa, hva du lovet, hvem du møtte og hva du burde følge opp.

AI-briller gjør dette mer interessant. De flytter da denne agenten fra en statisk skjerm og ut i møte med verden. Plutselig får agenten øyne, ører og en plass på kroppen som gjør at den kan samle fysisk data. Den ligger ikke bare i skyen eller på telefonen. Den sitter på nesen din og ser noe av det du ser, hører noe av det du hører og kan svare deg med stemmen. Du styrer agenten og interagerer med stemmen og ikke fingrene.

Da smelter flere av våre fire agenter sammen. Brillene er delvis en fysisk agent, fordi de har kamera, mikrofon og sensorer. De er en digital agent, fordi AI tolker og bearbeider informasjonen. De er en taleagent, fordi grensesnittet er naturlig samtale. Og de kan kobles til selvkjørende og fysiske systemer rundt deg, fra bilen til hjemme-roboten.

I denne sammenhengen blir brillene spesielt spennende som minnebank. Ikke nødvendigvis ved å lagre alt som rå video eller lyd. Det ville vært tungt og dyrt, og mest sannsynlig noe vi ikke ønsker... Men ved å lagre sammendrag, mønstre, nøkkelpunkter og meningsfulle spor. Hvem du møtte. Hva du liker. Hvilke steder du besøkte, også videre.

Da får vi noe som ligner en ekstern episodisk hukommelse. Du kan spørre brillene hva som egentlig ble sagt i møtet forrige uke. Hvilken restaurant dere var på i Tokyo. Hva navnet var på personen du traff på konferansen. Hva du tenkte da du stod foran et

bestemt objekt, bygg eller kunstverk. Ting som før forsvant i hjerneteppe, kan hentes frem igjen med stemmen.

En agent som kjenner historikken din, kan begynne å se mønstre du selv overser. Det kan forsterke intuisjonen vår. Den kan minne deg på at du ofte tar dårlige beslutninger når du er stresset. At du pleier å bli skeptisk til selskaper med visse typer språk i presentasjonene. At du sover dårligere etter bestemte møter, reiser eller arbeidsperioder. Det er ikke en erstatning for menneskelig mageføelse, men en datadrevet støtte til den.

Bevisstheten utvides også litt. Fordi de kan utvide det du selv legger merke til, som for eksempel å oversette skilt mens du går, forklare hva du ser på eller hente frem kontekst fra tidligere. Den gir et ekstra lag med informasjon.

Det betyr at minne blir en bro mellom agentene. Digitale agenter blir bedre når de får personlig og vedvarende kontekst. Taleagenter blir mer relasjonelle når de husker samtalen over tid. Fysiske agenter blir mer nyttige når de vet hva du pleier å gjøre og hvordan du vil ha ting. Selvkjørende agenter blir mer personlige når de forstår rutiner, preferanser og intensjoner.

Men alt dette skaper også nye knappheter. Batteri blir en flaskehals. Personvern må løses på en god måte uten å drepe innovasjonskraft. Så hvem eier minnene dine når de ligger i et system bygget av andre? Hva skjer når brillene husker bedre enn deg selv? Og hva gjør det med oss hvis vi slutter å stole på egen hukommelse fordi agenten alltid har fasiten?

Det er jo litt ubehagelig, men også utrolig spennende. For første gang kan digitale agenter bli en reell del av vårt kognitive system. Ikke bare et verktøy vi åpner, bruker og lukker. En digital kompis som følger oss gjennom dagen og gjør oss til supermennesker.

Kort sagt er AI-briller som daglig minnebank ikke bare en gadget. De kan bli det første praktiske steget mot en symbiotisk loop mellom menneske og agent. Vi går fra

bruker og maskin til menneske med ekstern hukommelse, forsterket intuisjone, samt en bevissthet som får teknologisk sidesyn.

En slags utvidet hippocampus på nesen. Litt rart. Litt skummelt. Men sannsynligvis en ganske stor greie.

En knapp digital ressurs

HBM er blitt avgjørende fordi moderne AI-modeller ikke bare trenger regnekraft. De trenger å flytte enorme mengder data raskt nok. Når kontekstvinduene vokser mot millioner av tokens, øker også behovet for KV-cache og aktivt minne kraftig. Minne er derfor blitt en strategisk ressurs.

Samtidig har digitale agenter et paradoks. De kan virke smarte i øyeblikket, men mange av dem er glemsomme over tid. De mangler ekte episodisk kontinuitet. De husker ikke livet ditt, relasjonene dine eller historikken din med mindre vi bygger egne lag for persistent minne rundt dem.

Her blir nettopp disse AI-brillene spennende. Da får vi noe som ligner en ekstern episodisk hukommelse. Du kan spørre hva som ble sagt i møtet sist torsdag, hvem personen på konferansen var, eller hva du tenkte da du stod foran et bestemt produkt. Det som før forsvant i tåken, kan hentes frem igjen.

Dette utvider også bevisstheten i øyeblikket. Brillene kan oversette, forklare, minne deg på ting og hente relevant kontekst uten at du må ned i telefonen. Intuisjonen kan også forsterkes når agenten ser mønstre i historikken din som du selv overser.

Slik oppstår en slags ny form for symbiotisk kognisjon. Mennesket kommer med mening, erfaring og følelser. Agenten kommer med lagring, søk og mønstergjenkjenning. Sammen blir de noe annet enn hver for seg.

Men spenningene er tydelige. Hvem eier minnet? Hva skjer når brillene husker bedre enn deg? Blir vi mer til stede fordi vi kan fokusere mer på å være i nuet, eller mindre til stede fordi vi outsourcer oppmerksomhet?

Dimensjon	Begrenset & Biologisk (mennesket)	Utvidet & Symbiotisk (AI briller & agenter
Minne	Rekonstruerende, feilbarlig, kapasitetsbegrenset (~7±2 i arbeidsminne), falmer over tid	Persistent, bit-nøyaktig på detaljer, episodisk over år, searchable via voice
Bevissthet	Smal spotlight, lett distrahert, krever aktiv oppmerksomhet	Utvidet spotlight: proaktive hint, kontekstualisert nåtid, redusert kognitiv last
Intuisjon	Rask, men upålitelig over tid, basert på implisitt minne	Hybrid: biologisk + datadrevet mønstergjenkjenning, kalibrert og kontekstualisert
Konsekvens	Autentisk, men begrenset og sårbar for glemsel & traume	Potensielt overflod av hukommelse, men risiko for avhengighet og "*ute av kroppen*"-følelse

De fire agentene frigjør oss fra mer enn arbeid. De flytter også grensen for hva teknologi kan tolke og utføre på våre vegne.

AI-briller som daglig minnebank er et tidlig tegn på dette. En digital agent tett på kroppen. En slags ekstern hippocampus som husker møtene, ansiktene, stedene og detaljene vi selv mister. Der menneskelig hukommelse glipper, kan agenten hente frem. Der vi rekonstruerer, kan den lagre mer presist.

Men presisjon er ikke det samme som mening. Menneskelig minne er feilbarlig. Det er påvirket av følelser, relasjoner og tid. Derfor er det også levende. Brillene kan hjelpe oss å huske mer, men de kan ikke velge hva som betyr noe for oss.

Spørsmålet er ikke om teknologien klarer det. Det gjør den gradvis. Det vi lurer på er hva som skjer når minnene våre flytter ut av hodet og inn i systemer bygget av andre.

Sakte, sakte, plutselig blir minne noe vi deler med maskinene.

I neste kapittel går vi mer inn i det tekniske.

2.0 Ulike minneteknologier

Periode	Dominerende teknologi	Nøkkelår & Milepæl	Kapasitet & Kostnad (grovt
1940–1950-tallet	Williams tube, Delay line, Magnetic drum	1947–1953 Første core memory (Whirlwind)	Bits → noen KB, veldig dyr per bit
1950–1970-tallet	Magnetic core memory	1953: MIT Whirlwind 1960-tallet: masseproduksjon	~100–1000 ns latency, dyr men pålitelig
1970-tallet	SRAM / DRAM introduksjon	1968–1970 Intel 1103 DRAM	Fallende kostnad, ~µs → ns latency
1980–1990-tallet	DRAM dominerer, Flash oppstår	1984 NOR Flash, 1987 NAND Flash (Toshiba	DRAM → billigere enn core; Flash → non-volatile
2000–2010-tallet	NAND eksploderer, SSD erstatter HDD	2000-tallet: NAND i mobil; 2010-tallet: 3D NAND	NAND → <$0.1/GB; DRAM ~$2–10/GB syklisk
2015–2020	HBM for GPU, Optane (PCM) prøver	2015+: HBM i GPU-er; Optane 2017–2022	HBM → høy båndbredde, men dyr
2023–2026	AI-drevet HBM-supercycle	2025–2026: HBM tar ~25% av DRAM-wafer	DRAM-priser +50–90 %, HBM ventetid 12–18 mnd

Minne pleide å være fysisk, dyrt og synlig. Store rom, magnetbånd og disker som duret. Så ble det gradvis noe vi sluttet å tenke noe særlig på. Fra rundt 2010 til 2022 føltes lagring nesten ubegrenset for vanlige brukere. Bilder, dokumenter, apper og sikkerhetskopier bare lå der i skyen. Billig nok til at minne sjelden var det første spørsmålet.

Når modeller skal trenes, kjøre inferens og holde lange kontekster aktive, er det ikke nok med rå regnekraft. Data må også flyttes raskt nok til at prosessorene ikke står og venter. Derfor har minne gått fra å være en commodity til å bli en av de viktigste flaskehalsene i hele AI-økonomien og en strukturell del av AI-infrastruktur utbyggingen.

Minne løser flere oppgaver samtidig. Det må gi lav forsinkelse, høy båndbredde og nok kapasitet til å holde data aktive. Det må enten kunne glemme når strømmen går, slik DRAM og HBM gjør, eller holde på data uten strøm, slik NAND, NOR, MRAM og ReRAM gjør. Og det må gjøre alt dette innenfor grenser for pris, strømforbruk og

plass. I AI betyr disse grensene mye. En modell kan være teoretisk mulig å kjøre, men økonomisk meningsløs hvis minne blir for dyrt eller energikrevende.

Epoke	Primær use case	Eksempler	Dominerende minne-type
Historisk (1980–2010)	Generell computing, PC, server	RAM til OS/apper, HDD/SSD til filer	DRAM + HDD → NAND SSD
Nåværende (2023–2026)	AI-trening & inference, edge AI, mobil	KV-cache i LLM-er, onboard robot/minne i briller	HBM (AI-GPU), LPDDR (mobil/edge), NAND (lagring)
Fremtidig (2027–2035)	Symbiotisk AI, universal memory, in-memory compute	Persistent kontekst i agenter, AI-briller som minnebank, CXL-minne pooling	HBM4/5 + HBF, MRAM/ReRAM embedded, FE-DRAM?

Derfor bygges datamaskiner som et minnehierarki. Nærmest prosessoren ligger SRAM, som brukes som cache. Det er ekstremt raskt, men tar for mye plass til å brukes i store mengder. DRAM og LPDDR fungerer som hovedminne i servere, PC-er og mobile enheter. HBM ligger tett på AI-akseleratoren og gir svært høy båndbredde, som er grunnen til at teknologien har blitt så sentral for AI-trening og inferens. HBM er ikke raskest på absolutt alt, men den kan flytte enorme datamengder parallelt. For store modeller er det ofte akkurat det som teller.

NAND og NOR har en annen rolle. NAND er billig per gigabyte og brukes i SSD-er, telefoner og datasentre der store datamengder må lagres. NOR brukes mer til kode og firmware der pålitelighet er viktig. De er tregere enn arbeidsminne, men de vinner på kapasitet og kostnad.

De nye minneteknologiene prøver å fylle gapet mellom raskt arbeidsminne og permanent lagring. MRAM, ReRAM, FeRAM og PCM kan holde på data uten strøm, men samtidig gi raskere tilgang enn tradisjonell lagring i enkelte bruksområder. De kommer neppe til å erstatte DRAM eller NAND over natten. Mer sannsynlig blir de viktige i edge-enheter, sensorer, roboter og AI-briller, der strømforbruk og rask tilgang til personlig kontekst betyr mer enn ren lagringsmengde.

AI-brillene derimot, kan ikke lene seg på datasenterets strømforbruk. De trenger minne som er lett, raskt og energieffektivt nok til å fungere på kroppen gjennom en hel dag. Hvis brillene skal huske møter, steder, samtaler og preferanser, må noe av minne flyttes nærmere brukeren. Ikke som rå videolagring av alt vi ser, men som komprimerte spor, sammendrag og embeddings.

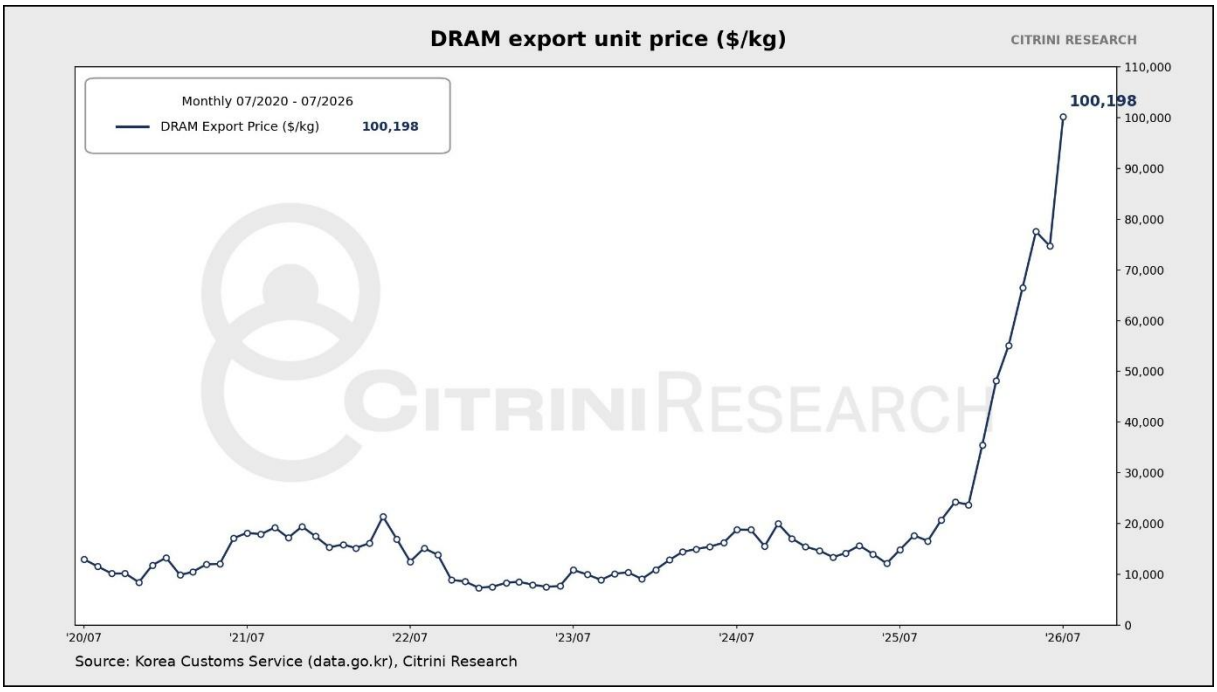
I datasentre er HBM den viktigste knappheten akkurat nå. Store språkmodeller og multimodale modeller trenger enorm båndbredde, og KV-cache gjør presset enda tydeligere når tidligere tokens må holdes aktive under inferens. Dette gjør minne til en direkte kostnadsdriver for AI.

I den gamle dataverdenen snakket vi mest om prosessorer. I den nåværende AI-verdenen må vi snakke like mye om minne. Det avgjør hvilke modeller som kan kjøres, hvor lange kontekster de kan håndtere, hva inferens koster, og hvor langt agentene kan flyttes ut fra skyen før batteri, latens og lagring setter grenser.

2.1 Hvorfor er det krise nå?

AI har gjort minne til en flaskehals. Store modeller krever mye aktivt minne, særlig når lange kontekster gir stor KV-cache. Derfor har HBM blitt kritisk i AI-infrastruktur. Uten nok båndbredde venter akseleratorene på data.

Samsung, SK hynix og Micron prioriterer HBM fordi AI-kundene betaler mest. HBM bruker langt mer wafer-areal per gigabyte enn vanlig DRAM. Når kapasitet flyttes dit, blir DRAM og deler av NAND-markedet strammere. Prisene stiger.



Dette bryter med den gamle minnelogikken. Historisk har høyere volum gitt lavere kostnad. Nå trekker AI kapasitet mot de dyreste produktene og presser resten av markedet opp. Wright's Law virker over tid, men akkurat nå vinner knapphet.

Teknologien flyttes nærmere beregningen, som vil si at HBM pakkes tett på prosessoren gjennom 2.5D og 3D-stacking. CXL gjør minne mer delbart på systemnivå. High Bandwidth Flash prøver å fylle gapet mellom DRAM og NAND. Målet er enkelt. Kortere vei for data, lavere energitap.

På edgen gjelder andre krav. AI-briller, sensorer og små roboter trenger lavt strømforbruk og *persistent state*. MRAM ligger lengst fremme. ReRAM går mot edge AI og neuromorfisk databehandling. FeRAM og FeFET passer der batteri er viktigst. PCM lever videre i mer spesialiserte nisjer.

Disse teknologiene løser andre problemer og erstatter ikke DRAM eller HBM med det første. Datasenteret trenger ekstrem båndbredde. Edge-enheter trenger minne som varer uten strøm. Begge trekker mot samme konklusjon. Minne avgjør hvor agentene kan kjøre, hvor lenge de kan huske, samt hva det vil koste å bruke dem.

Postcards From The Future

Selskaper & aktører i embedded og edge

	Lav **Modenhet & Volum-produksjon** masseproduksjon/tilgjengelig IP	**Høy** **Modenhet & Volum-produksjon** masseproduksjon/tilgjengelig IP
Høy **Edge AI & Lav-strøm-fokus**	Weebit Nano (ReRAM) Avalanche Technology (ReRAM/MRAM) Crossbar (ReRAM)	Samsung (14–8 nm eMRAM) TSMC (22 nm eMRAM/ReRAM) GlobalFoundries (22–12 nm eMRAM/ReRAM)
Lav **Edge AI & Lav-strøm-fokus**	Noen FeFET-startups PCM-spesialister (lavere edge-fokus)	Everspin (standalone MRAM) Fujitsu / Sony / Renesas (FeRAM) SK hynix (PCM / hybrid)

ent

Postcards From The Future

Use cases i embedded & edge

	Lav latency-krav (kan tåle µs–ms)	**Høy latency-krav** (<50–100 ns read/write)
Høyt strømforbruk Batterikrav høyt	**IoT sensors** (lang levetid, sjelden wake-up) Code storage / config data (FeRAM / ReRAM)	**Edge AI-briller, humanoids, autonome droner.** Edge AI inference (on-device weights, KV-cache lite) Always-on voice/vision (MRAM / ReRAM)
Lavt strømforbruk	**Industrial controllers** Automotive BMS / domain controllers (MRAM / FeRAM)	**5G Edge gateways & industri roboter** Real-time control loops (robotics, ADAS) Neuromorphic / analog compute edge (ReRAM / FeRAM)

ent

Emerging minne ser ut til å vinne i embedded og edge fordi det løser en annen oppgave enn HBM i datasenteret. Her er ikke målet å lagre enorme datamengder billigst mulig. Målet er å holde små mengder kritisk informasjon tett på enheten, uten å bruke mye strøm og uten å miste alt når strømmen kuttes. For en AI-brille, en

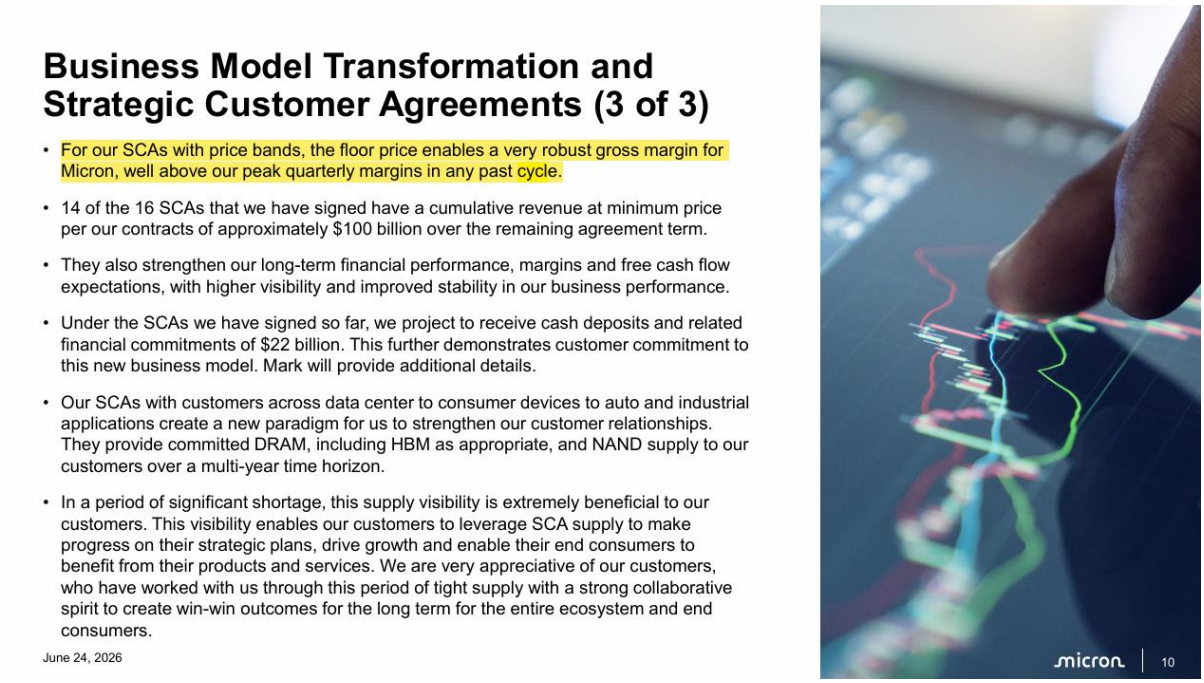
sensor eller en liten robot kan 1 til 100 MB persistent minne være mer verdifullt enn store mengder treg lagring et annet sted i systemet.

Dette er grunnen til at MRAM, ReRAM, FeRAM og PCM får mer oppmerksomhet. De gir ikke samme tetthet som NAND, og erstatter ikke HBM i tunge AI-systemer. Men de kan lagre tilstand, modellvekter og lokale logger direkte i chipen. Det reduserer behovet for eksterne minnebrikker, kutter latensen og sparer areal i små systemer. MRAM leder på modenhet og foundry-støtte. ReRAM er mest spennende for analog edge AI og in-memory compute. FeRAM passer best der strømforbruket må helt ned. PCM blir mer nisje, men kan fortsatt være nyttig i miljøer med høye krav til temperatur og robusthet.

Fra commodity til strategisk ressurs

Historisk har minne vært en av halvlederindustriens mest brutale sykliske markeder. Når prisene steg, bygget produsentene kapasitet. Når kapasiteten kom på plass (oversupply), falt prisene igjen. Kundene kunne vente. Produsentene hadde begrenset prisingsmakt. Så begynte AI å endre dynamikken.

Micron har nå inngått 16 såkalte Strategic Customer Agreements. Selskapet forventer at disse etter hvert vil dekke rundt halvparten av omsetningen. Mange løper helt til 2030 og inneholder bindende *take-or-pay*-mekanismer. Kunden reserverer dermed ikke bare en komponent. Den sikrer seg tilgang til en ressurs den frykter å mangle.



Det er et viktig skifte. For en hyperscaler er kostnaden ved å betale litt for mye for minne relativt liten. Kostnaden ved å ha milliarder av dollar i GPU-er stående uten nok minne kan være enorm. Når konsekvensen av mangel blir større enn konsekvensen av høy pris, endres forhandlingsmakten.

Micron kommuniserer:

«In the AI era, memory has become a strategic asset for our customers"

Dette bidrar til å forklare hvorfor denne oppgangen kan få en annen karakter enn tidligere sykluser. Produsentene bygger fortsatt fabrikker og kapasitet, men stadig mer etterspørsel kontraktsfestes før kapasiteten eksisterer. Minne går fra spotmarked mot en struktur som på enkelte områder likner mer om energi, foundry-kapasitet eller langsiktige industrikontrakter.

Peak-cycle-risikoen forsvinner derimot ikke. Men dersom produsentene får bedre etterspørselsvisibilitet, prisgulv og flerårige volumforpliktelser, kan både marginer og investeringsbeslutninger bli mindre volatile enn det historisk normalt sett bruker å bli.

Hvilke lover kan hjelpe oss å forstå bedre?

Bak dette ligger flere kjente teknologilover som nå peker i samme retning. Moore's Law ga oss lenge flere transistorer på samme areal, men den effekten er blitt dyrere å hente ut. Rock's Law gjør det enda tøffere, fordi nye avanserte fabrikker krever investeringer på titalls milliarder dollar. Wright's Law har historisk presset kostnadene ned når produksjonsvolumet steg, særlig for DRAM og NAND. I 2026 ser vi en merkelig pause i den logikken. AI trekker produksjonskapasitet mot HBM, og vanlige minnetyper blir dyrere før ny kapasitet rekker å komme på plass.

Den dypere begrensningen er den såkalte «Memory Wall». Prosessorene blir kraftigere, men data flytter seg ikke raskt nok. I store AI-modeller blir dette synlig gjennom KV-cache og lange kontekstvinduer. Når en modell må holde enorme mengder informasjon aktivt under inferens, blir minne en direkte kostnadsdriver. Amdahl's Law trekker mot samme problem. Når én del av systemet akselererer kraftig, flyttes flaskehalsen til den tregeste delen. I AI er det ofte minne.

Koomey's Law har vist at beregning blir mer energieffektiv over tid, men stadig mer strøm brukes på å flytte data frem og tilbake. Derfor vokser interessen for *near-memory* og *in-memory compute*. Jo kortere data må reise, desto mindre energi forsvinner på transport i stedet for arbeid.

Vi står derfor midt i en minne-supersyklus. HBM prioriteres i datasenteret, vanlig DRAM og deler av NAND-markedet strammes til, og prisene presses opp. Flaskehalsen kan vare til 2027 eller 2028, avhengig av ny kapasitet, HBM4 og løsninger som HBF. Samtidig vokser et annet minnemarked frem på kanten, der persistent state og lavt strømforbruk betyr mer enn bare rå båndbredde.

Fremtiden blir nok et mer sammensatt minnehierarki hvor HBM tar de tyngste AI-jobbene og NVM får en viktigere rolle i briller, roboter og små edge-systemer. Arkitekturen bygges om for å flytte minne nærmere beregningen. Det markerer

slutten på den enkle skaleringen der flere transistorer alene løste problemet. Nå handler fremskritt like mye om hvordan data lagres, flyttes og hentes frem.

3.0 In-memory computing: Nøkkelen for edge-agenter

In-memory computing, ofte forkortet IMC eller PIM, betyr at beregningen flyttes inn i eller svært tett på minne. I stedet for at data hele tiden sendes frem og tilbake mellom minne og prosessor, gjøres deler av arbeidet der dataene allerede ligger.

Det bryter med den klassiske von Neumann-arkitekturen, som fortsatt ligger under det meste av moderne databehandling. Der er minne og prosessor skilt fra hverandre. Minne lagrer. Prosessoren henter, regner og skriver tilbake. Modellen har fungert i mange tiår, men AI får frem svakheten. Ved store datamengder, bruker systemet stadig mer tid og energi på transport av data.

Dette er kjernen i utfordringen bak nevnte memory wall. Prosessorene blir kraftigere, men dataflyten henger ikke med. I datasentre blir det dyrt. I edge'n er det kritisk. En AI-brille eller en drone har ikke ubegrenset batteri eller plass. Hvis mesteparten av energien går med til å flytte data, stopper skaleringskurven fort.

IMC forsøker å kutte denne sløsingen. I digital IMC gjøres enkle operasjoner eller matriseberegninger direkte i minnearrayet (minnecellene der dataene lagres) eller tett ved det. SRAM-baserte løsninger brukes særlig der presisjon og moden produksjon teller. Analog IMC går lenger. Her utnyttes fysiske egenskaper i teknologier som ReRAM, MRAM, PCM eller FeFET for å gjøre vektorberegninger mer energieffektivt. Det kan gi store gevinster i strømforbruk, men presisjon og støy blir vanskeligere å kontrollere.

Near-memory compute er en litt mer praktisk mellomløsning hvor beregningen ikke skjer inne i selve minnecellen, men nær nok til at dataflyttingen reduseres kraftig. HBM med logikklag og avansert pakking driver mot denne retningen. Det samme gjør CXL-baserte arkitekturer, der minne kan deles mer fleksibelt mellom systemer.

For edge-agenter er dette spesielt viktig. En agent på kroppen eller i en liten maskin må tolke sensorinput, holde på lokal kontekst og svare raskt. Den kan ikke sende alt til skyen hver gang. Latensen blir for høy og batteriet tappes. Mer beregning må derfor skje lokalt, og da blir minne en del av selve regnemotoren.

Hvorfor er det så sentralt for våre agenter som lever i edgen?

IMC blir sentralt fordi agentene våre må virke der verden faktisk skjer. På kroppen, i maskinen og ute i miljøet. De kan ikke vente på et datasenter hver gang de skal tolke en lyd, forstå et objekt eller hente frem personlig kontekst.

Problemet er at edge-enheter har langt mindre å gå på enn datasentre. I vanlig arkitektur flyttes data hele tiden mellom minne og prosessor. Det koster tid. Enda verre, det koster strøm. I flere edge-workloads er dataflytting den største energislukeren.

For en digital agent i skyen er dette dyrt, men det er eksistensielt for våre fysiske agenter ute i edge'n . En humanoid som skal gripe et objekt må reagere raskt. En selvkjørende bil kan ikke vente på en rundtur til skyen hvis forbindelsen hakker.

IMC angriper flaskehalsen ved roten. Når deler av beregningen skjer i eller nær minne, slipper systemet å flytte like mye data frem og tilbake. Det gir lavere latens og bedre batteritid. For edge-agenter kan det være forskjellen på en demo som varer i 40 minutter og et produkt som faktisk kan brukes gjennom en arbeidsdag.

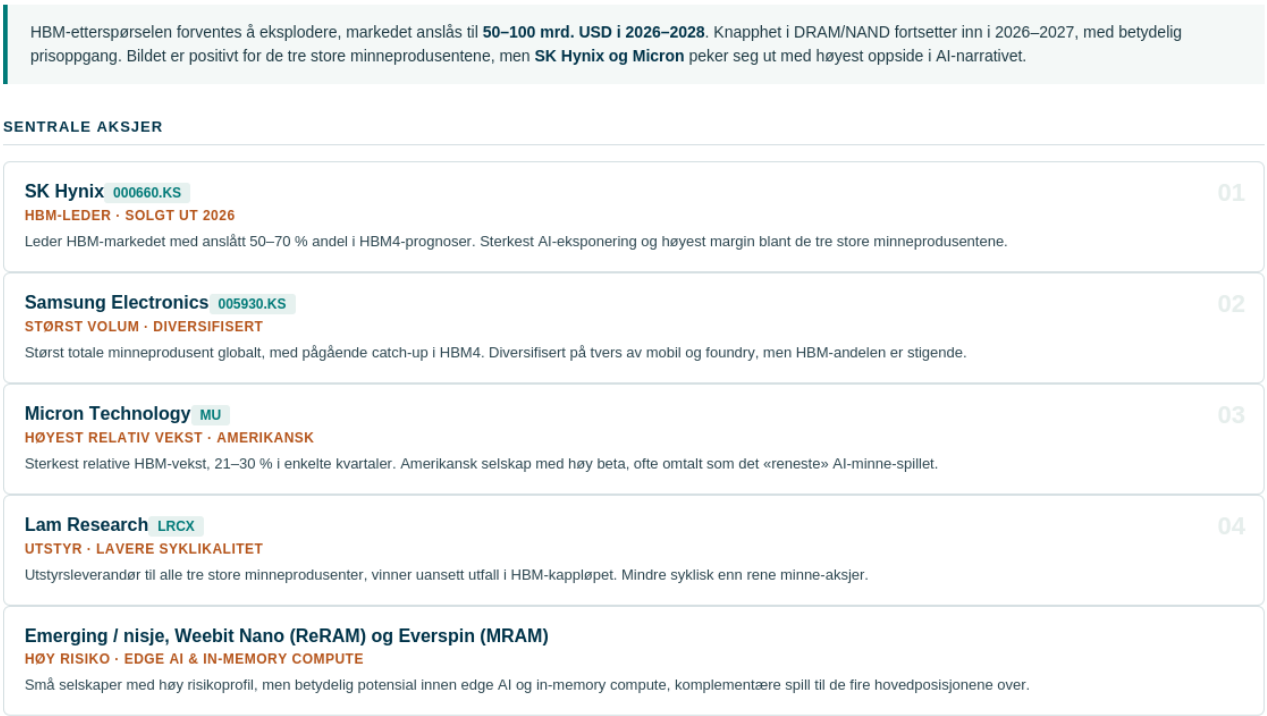
Persistent state gjør dette enda viktigere. Agentene trenger å huske preferanser, tidligere samtaler og lokalt miljø uten å bruke strøm på konstant refresh. DRAM er

raskt, men glemmer uten strøm og må holdes aktivt. Non-volatile IMC med teknologier som MRAM, ReRAM eller FeFET kan holde på tilstand mer effektivt. Det passer bedre for alltid-på-enheter som skal våkne raskt og fortsette der de slapp.

Dette er også grunnen til at emerging minne og IMC bør sees sammen. MRAM og ReRAM er ikke bare nye lagringsformer. De kan bli byggesteiner i arkitekturer der lagring og beregning glir mer sammen. For edge AI er det attraktivt fordi areal, batteri og responstid er de harde begrensningene.

Uten IMC blir edge-agentene for avhengige av skyen. De blir tregere, mer strømkrevende og mer sårbare for dårlig tilkobling. Med IMC kan mer intelligens flyttes lokalt. Agenten kan reagere raskere, holde bedre på kontekst og sende mindre sensitiv data ut av enheten.

4.0 Sentrale minneaksjer



Kilde: Bank of America

To av de mest fremtredende i SMB-markedet

Weebit Nano og Everspin Technologies fokuserer på å erstatte eller supplere tradisjonell embedded flash (eFlash/NOR) med bedre ytelse, lavere strømforbruk og høyere endurance, noe som er spesielt interessant for edge AI-agenter, IoT, bil og batteridrevne enheter.

Weebit Nano (ASX: WBT) - ReRAM-leder for embedded edge AI

Weebit Nano er en australsk-basert IP-leverandør (ikke full produksjon, men lisensierer ReRAM-teknologi til foundries og IDM'er). De har hatt sterk momentum i 2025-2026, med flere kommersielle milepæler.

ReRAM er produksjonsklar i flere noder (f.eks 22 nm hos DB HiTek og onsemi). De har oppnådd JEDEC-kvalifisering hos partnere og viser live edge AI-demoer (f.eks. på Embedded World 2026 og CES-messen som var nå i 2026).

STORE PARTNERSKAP

Texas Instruments DES. 2025 / JAN. 2026	Tier-1-lisensavtale for embedded prosessorer og produkter. Det viktigste kommersielle valideringspunktet for teknologien så langt.
DB HiTek (Korea) MARS 2026	Teknologikvalifisert og valgt til nasjonalt Compute-in-Memory-program for AI in-memory compute. Gir Weebit fotfeste i det koreanske halvlederøkosystemet.
onsemi TAPE-OUT	Tape-out av testchips med fokus på ultra-low-power edge-anvendelser. Ytterligere bekreftelse på teknologiens egnethet for strømkritiske design.

FORDELER I EDGE

Teknologien bruker lite strøm når data skrives, og tåler langt flere skriveoperasjoner enn tradisjonelt flash-minne uten å slites ut. Det gjør den godt egnet for AI-systemer som hele tiden oppdaterer og lagrer informasjon, for eksempel AI-agenter som må huske kontekst over tid.

RISIKO OG UTSIKTER

Fortsatt tidlig i volum. Weebits IP-modell gjør selskapet avhengig av partnernes egne lanseringer og opptak av teknologien. 2026 trekkes frem som et potensielt vendepunkt, med flere avtaler ventet i tillegg til de tre allerede inngått.

Everspin Technologies (MRAM US) - Etablert MRAM-spiller med fokus på high-reliability

Everspin er den mest modne kommersielle MRAM-aktøren (verdensledende i standalone og delvis embedded MRAM). De har lengre track record enn de fleste emerging-spillere.

Everspin leverer STT-MRAM med fokus på pålitelighet, robusthet og lang levetid i krevende miljøer, med sterk posisjon i aerospace, forsvar, bil og industriell IoT.

TEKNOLOGISTATUS

PERSYST STT-MRAM-familien (xSPI-grensesnitt) er i høy pålitelighet-segmentet (HR). 64 Mb er kvalifisert (AEC-Q100 Grade 1, mars 2026), 128 Mb-kvalifisering kommer mai 2026 og 256 Mb i juli 2026, med volumproduksjon ventet andre halvår 2026.

NYE PRODUKTER

UNISYST MRAM (lansert mars 2026) er et samlet minne, kode og data i én ikke-flyktig arkitektur, rettet mot edge AI, industrielle formål og oppgaver der feil ikke kan aksepteres. I tillegg er EMxxLX-serien utvidet med høyere tetthet og et bredere temperaturområde, fra −40°C til +125°C.

APPLIKASJONER I EDGE OG EMBEDDED

Sterk posisjon i aerospace, forsvar, bilindustri (ECU-er), industriell IoT, satellitter (LEO) og robotikk. Validert for konfigurasjon av Lattice FPGA-er, med vekt på datalagring over tid, dataintegritet, strålingstoleranse og umiddelbar oppstart uten boot-tid.

FORDELER

Svært høy utholdenhet (over 10^{14} skrivesykluser), rask (10–50 ns) og ikke-flyktig uten behov for oppdatering. Robust i ekstreme miljøer. Mindre rettet mot analog in-memory compute enn ReRAM, og mer mot digitale løsninger med høy pålitelighet.

RISIKO OG UTSIKTER

Mer etablert enn typiske oppstartsselskaper, med lavere beta, men fortsatt volatil i takt med minnesyklusene. Selskapet står sterkt i nisjer som aerospace og industri, men er et mindre rendyrket AI-spill enn Weebit.

Sammenlikning per H1 2026

	Lav volumproduksjon (tidlig)	**Høy volumproduksjon (moden)**
Høy edge-fokus	**Weebit Nano** ReRAM IP og lisensmodell Nærmer seg volumproduksjon, men ikke helt der ennå	
Lav edge-fokus		**Everspin** Standalone og høyt forutsigbar MRAM Fokus på bil- og flybransjen

Weebit er det mer hype-drevne, høy-beta-spillet med sterkest edge AI-potensial (ReRAMs analog compute-fordel + lisensvekst). 2026 kan bli breakout-år.

Everspin er det mer stabile, etablerte valget. De er allerede i produksjon, sterk i mission-critical edge (ikke like AI-intensiv, men robust).

Begge drar nytte av memory wall og edge AI-trenden i *de fire fytternes renessanse*, men Weebit har høyere oppside/risiko i AI-briller/humanoids, mens Everspin vinner i pålitelige industri-sammenhenger.

Vår porteføljeposisjonering

Stock selection and portfolio management

Case study: Memory - identifying the AI bottleneck

Thesis highlights

What we identified

Memory is a critical bottleneck for AI training and inference, not just a commodity input.

Our thesis

Structural supercycle, not a normal memory cycle: HBM demand scales with AI compute, supply is consolidated across a handful of players.

What we did

Built positions across Micron, SK Hynix, Samsung and Sandisk with active sizing across the names on risk/reward and cycle exposure.

Ongoing active management

Sandisk trimmed (higher risk NAND exposure). Micron is largest (most upside). Considering rotation from Micron to SK Hynix as relative valuation shifts.

Current positioning in memory

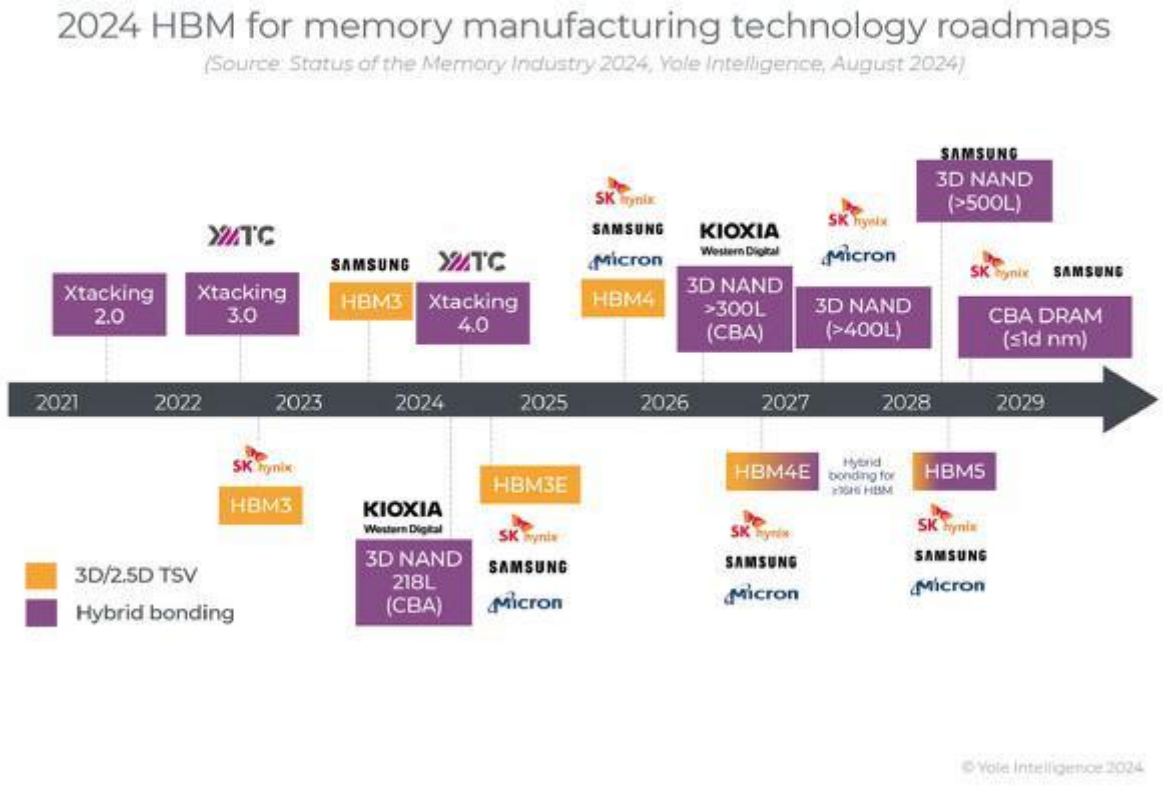
Micron 4.7%
Largest position - still upside to DRAM prices

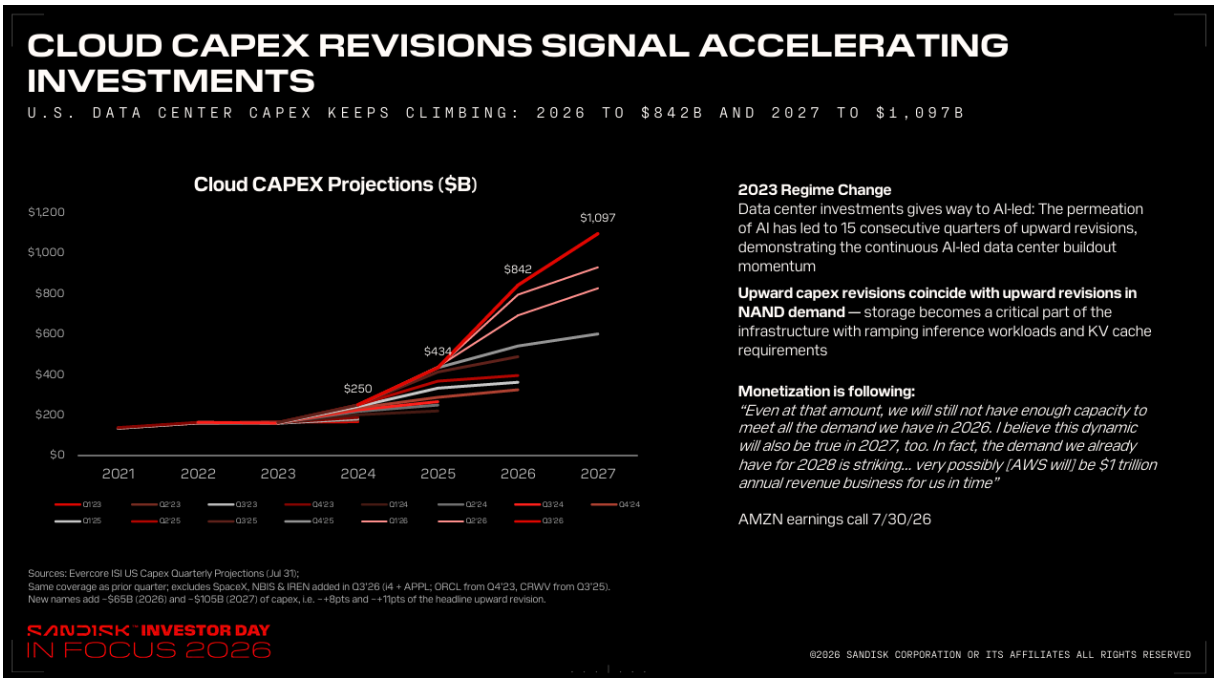
Samsung 2.1%
Diversified exposure - still attractive valuation

Sandisk 1.8%
Trimmed on outperformance and higher risk NAND profile

SK Hynix 1.3%
HMB leader - lower disruption risk, but less upside from pricing

5.0 Roadmaps og utvikling





Kilde: Sandisk Investor Day 2026

5.1 HBF

Vi har hittil i dette perspektivnotatet beskrevet minnehierarkiet som relativt ryddig. HBM og DRAM ligger tett på prosessoren fordi de er raske. NAND ligger lenger unna fordi det er billigere og kan lagre langt mer informasjon.

Nå forsøker Sandisk å gjøre grensen mindre tydelig. High Bandwidth Flash, eller HBF, bygger på NAND, men pakkes og kobles sammen for å levere mye høyere parallell båndbredde enn tradisjonell NAND flash. Målet er ikke nødvendigvis å erstatte HBM. Det er å fylle et voksende hull mellom ekstremt raskt, dyrt DRAM-basert minne og svært kapasitetsrikt NAND.

Sandisk beskriver utfordringen fra et inferensperspektiv. Store modeller trenger både båndbredde og stadig mer kapasitet nær prosessoren. HBM er svært raskt, men kapasiteten er dyr. NAND har enorm kapasitet, men har tradisjonelt vært for tregt. HBF forsøker å kombinere deler av begge verdener.

Sandisk hevder at teknologien kan levere HBM-lignende båndbredde samtidig som den gir rundt 8 til 16 ganger høyere kapasitet enn HBM ved sammenlignbar kostnad.

I Mizuho-notatet etter selskapets investordag beskrives en mulig 16-stack HBF-løsning som HBM-lignende båndbredde til rundt en tidel av kostnaden, med første dies (brikkene) og controllere (styringsenheter) mot slutten av 2026 eller starten av 2027 og mulig inntekstsbidrag fra 2028. Spørsmålet blir dermed ikke «HBM eller HBF?». Det kan bli begge deler.

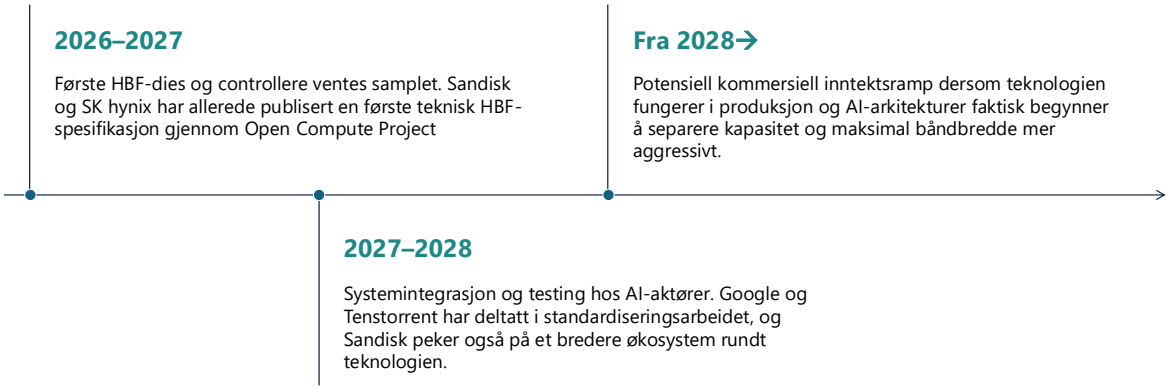
HBM holder de mest latens-sensitive dataene tett på beregningen. HBF kan fungere som et langt større lag med modellvekter eller annen informasjon litt lenger ute. I stedet for å betale HBM-pris for hver byte modellen potensielt trenger, kan systemet bygge et mer differensiert minnehierarki.

Det ligner litt på et skrivebord. HBM er dokumentene du har åpne foran deg. HBF kan være arkivskapet rett borti hjørnet. NAND-SSD er lageret nede i kjelleren.

Poenget er ikke å gjøre arkivskapet like raskt som hånden din. Det holder at du slipper å løpe ned i kjelleren hver gang du trenger noe.

For inferens kan nettopp dette bli viktig når modeller, kontekstvinduer og agentminne vokser raskere enn det er økonomisk mulig å fylle systemene med HBM.

High Bandwith Flash: Roadmap



Den store opsjonen vil kunne være at HBF åpner for at NAND tar en større del av AI-minneverdikjeden enn klassiske SSD-er alene ville gjort.

Flere lag av memory wall

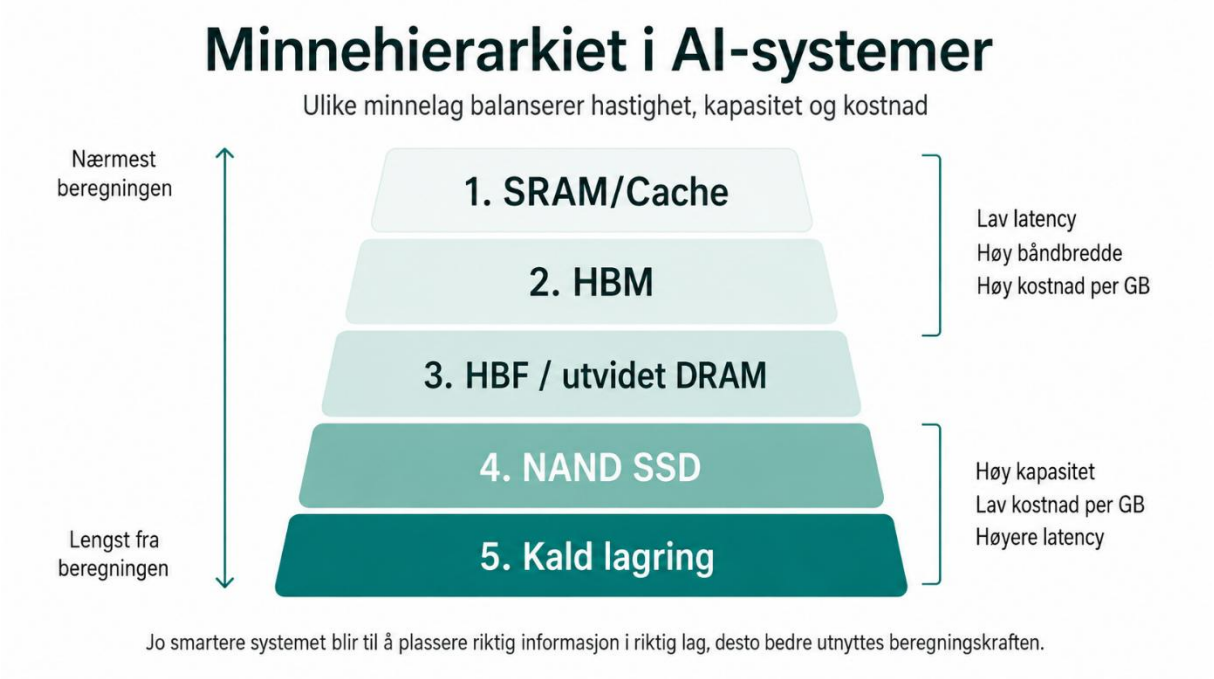
Memory Wall omtales ofte som ett problem. I virkeligheten finnes det flere vegger.

Den første er båndbreddeveggen. GPU-en klarer å regne raskere enn minnet klarer å mate den med data. HBM forsøker å løse denne.

Den neste er kapasitetsveggen. Selv om HBM er ekstremt raskt, blir det økonomisk og fysisk vanskelig å plassere stadig større mengder HBM rundt hver akselerator. Her kan teknologier som HBF, CXL og større LPDRAM-baserte minnelag bli viktige.

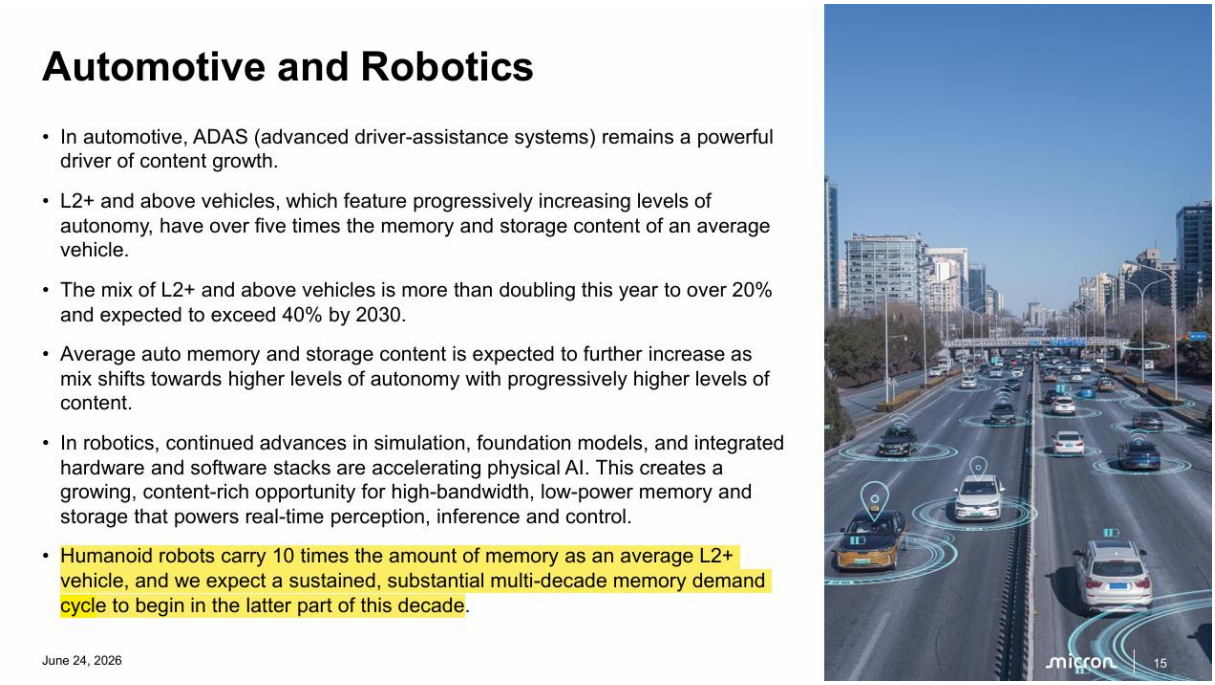
Så kommer energiveggen. Data som flyttes frem og tilbake bruker strøm. Jo større AI-systemet blir, desto mer energi går til å flytte informasjon og ikke bare til selve beregningen. Det er dette som gjør in-memory og near-memory computing mer og mer interessant.

AI løses derfor neppe av én magisk minneteknologi. Resultatet kan bli et stadig dypere hierarki:



Jo smartere systemet blir til å plassere riktig informasjon i riktig lag, desto bedre utnyttes den dyreste beregningskraften. Memory Wall forsvinner altså ikke nødvendigvis. Vi bygger heller flere åpninger gjennom den.

5.2 Hvor mye minne trenger egentlig en humanoid?



Kilde: Micron Q2 2026 Report

Datasenteret er bare den ene siden av minnehistorien.

Den andre handler om hva som skjer når kunstig intelligens kommer ut av treningssentrene (datasentre) og ut av skjermene og inn i den virkelige fysiske verden.

En humanoid robot er i praksis en mobil AI-maskin med en krevende utfordring. Den må kontinuerlig se verden, tolke tale, forstå rommet rundt seg, planlegge bevegelser og kontrollere et stort antall motorer. Alt dette skjer samtidig, og mye av det må skje lokalt.

I et datasenter kan en modell vente noen millisekunder ekstra eller hente informasjon fra en annen server. En humanoid som er på vei til å miste balansen har ikke den luksusen.

Det skaper flere typer minnebehov samtidig. Roboten trenger raskt arbeidsminne for persepsjon og planlegging. Den trenger lokal lagring for modeller og sensorhistorikk. Den kan trenge persistent minne for å huske kalibrering, miljø og tilstand uten at alt må lastes inn på nytt hver gang den starter. Etter hvert som humanoiden blir mer personlig, får den i tillegg et nytt lag: episodisk agentminne.

Det siste er interessant. En industrirobot trenger hovedsakelig å huske oppgaven. En generell humanoid kan måtte huske verden den opererer i.

Hvor ligger verktøyene? Hvordan foretrekker denne brukeren at kjøkkenet organiseres? Hvilke objekter er skjøre? Hva skjedde forrige gang den prøvde denne oppgaven?

Da går minne fra å være hardwarekapasitet til å bli en del av robotens læring over tid. Jo mer autonom roboten blir, desto større blir verdien av lokal kontekst.

Hvor minne blir mest kritisk

	Lav lokal autonomi	**Høy lokal autonomi**
Høyt sensorvolum	Cloud-assistert kamera	Humanoid / selvkjørende bil (mest kritisk)
Lavt sensorvolum	Enkel IoT / sensor	AI-briller / wearables

6.0 Avslutning

Minne blir som en bro mellom hjernen og maskinen, mellom det vi opplever og det teknologien etter hvert kan lagre for oss.

Menneskelig minne er aldri helt presist. Vi henter ikke frem det vi husker alltid helt presist som det skjedde. Vi bygger det opp igjen, hver gang, med nye følelser og ny kontekst. Det gjør oss upålitelige som arkiv, men sterke som mennesker. Vi kan se mening i det ufullstendige. Vi kan glemme litt for å leve videre.

Teknologisk minne går motsatt vei. Det vil være raskere, billigere per oppgave og mer tilgjengelig. HBM løfter datasenteret. Emerging minne flytter tilstand ut i edge-enheter. In-memory computing forsøker å kutte selve sløsingen i dataflyttingen. Alt peker mot samme utvikling. Agentene skal huske mer lokalt og bli mindre avhengige av skyen.

Det er her AI-brillene blir et godt bilde på overgangen, fordi de plasserer agenten tett på dataen. De kan huske møtet du glemte, navnet som glapp, eller konteksten du ikke rakk å hente frem. Etter hvert blir de en personlig minnebank, ikke som en harddisk på nesen, men som et lag av små spor som følger deg gjennom dagen.

Resultatet er en ny form for kognitiv symbiose. Mennesket bidrar med erfaring og mening. Maskinene med presis lagring og mønstergjenkjenning. Sammen kan det vi legger merke til utvides, og gi intuisjonen et bedre datagrunnlag. Det kan gjøre oss klokere, men også mer avhengige.

En stor risiko er at vi slutter å eie forholdet til våre egne minner. Hvis samtaler og mennesker blir søkbare gjennom systemer bygget av andre, flyttes også makt. Spørsmålet om minne blir da et spørsmål om eierskap og tillit.

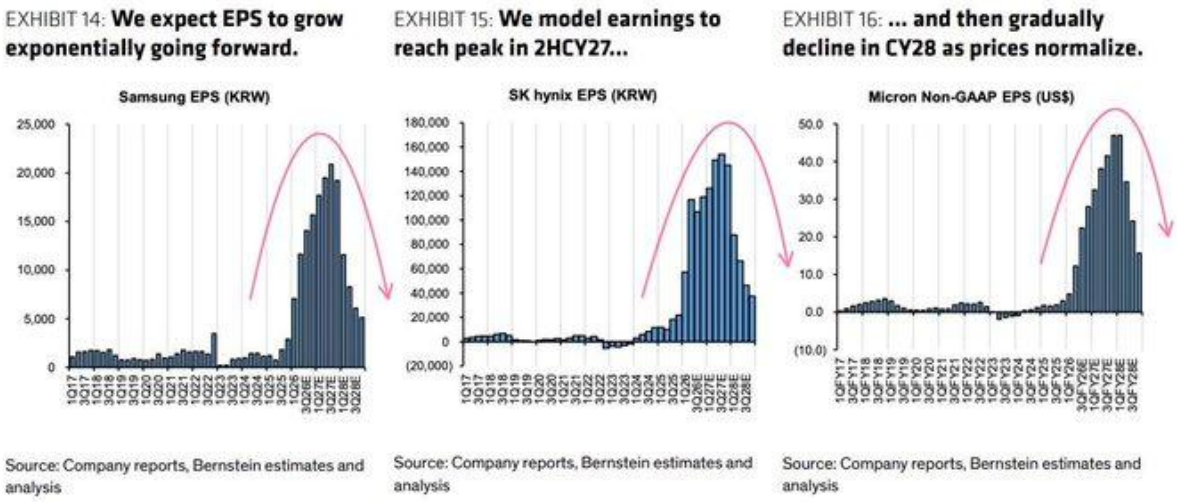
De fire agentenes renessanse frigjør oss fra mye arbeid. Kanskje også fra noe av vår egen kognitive begrensning. Men den bør ikke frigjøre oss fra ansvaret for å velge

hva som betyr noe. Maskinen kan lagre nesten alt. Den kan ikke vite hva som bør forme et liv.

Sakte, sakte, plutselig, blir minne noe vi deler med teknologien. Det trenger ikke være et tap av menneskelig bevissthet og hukommelse. Det kan bli en dypere form for hukommelse, hvis vi bygger den riktig og fortsatt lar mennesket bestemme hva som er verdt å huske.

6.1 AI-supercycle eller klassisk peak cycle?

I år rapporterer Micron ekstremt sterk etterspørsel etter minne. Kundene får bare rundt 60% av det de bestiller. Samsung og SK Hynix inngår nå lengre 3–5 års kontrakter, og supplymangelen forventes å vare til 2030. Likevel faller aksjen. Markedet priser inn en klassisk *peak cycle* og tror at E'en i P/E skal ned. Det kan beskrives som toppen av en syklus der boomen snart snur til bust. Men er det riktig?



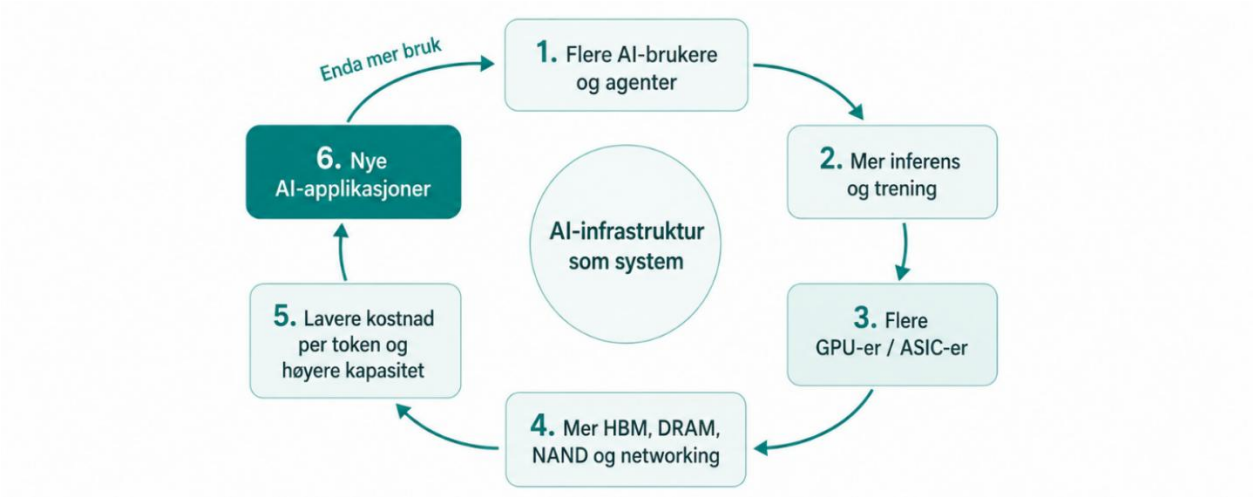
Hva hvis det ikke blir tilfellet?

Er vi midt i en ny, strukturell AI-*supercycle* drevet av teknologi som bryter *Memory Wall?* Denne delen av perspektivnotatet ser nærmere på denne problemstillingen. Fra definisjoner via matriser og analogier til de konkrete teknologiene som fungerer som enablere.

6.2 AI-flywheel

Det er jo lett å tenke på AI-infrastruktur som en serie separate markeder. GPU-er, minne, networking, datasentre også videre. Men de oppfører seg stadig mer som ett system.

Jensen Huang har nylig anslått at halvlederindustrien kan måtte bli mange ganger større det neste tiåret. Begrunnelsen er den samme som går igjen gjennom hele AI-infrastrukturen. Mer intelligens krever mer beregning. Mer beregning krever mer minne og nettverk. Når denne infrastrukturen bygges, blir AI billigere og bedre. Det skaper igjen flere anvendelser som krever enda mer infrastruktur.

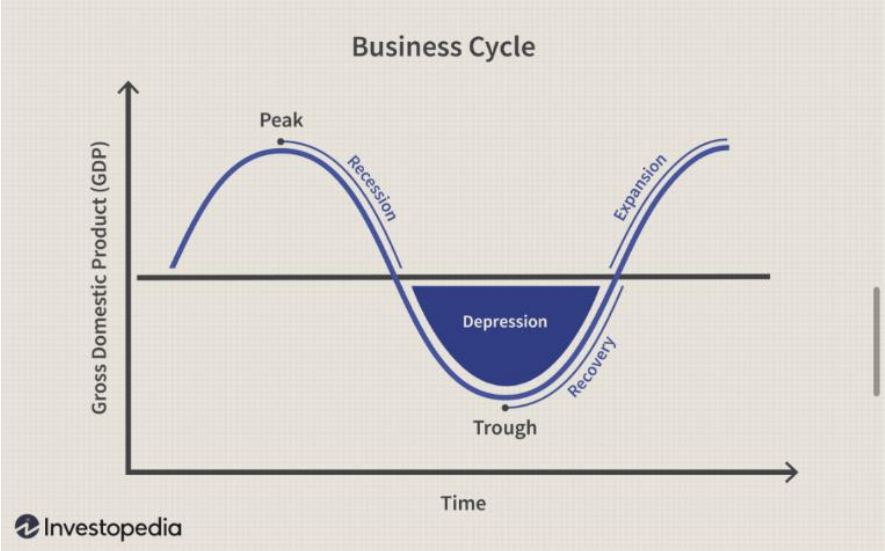


Dette er årsaken til at dagens knapphet kan være vanskeligere å løse enn en vanlig vareknapphet. Ny supply møter ikke nødvendigvis den gamle etterspørselen. Den møter en etterspørsel som selv har rukket å vokse fordi mer kapasitet ble tilgjengelig. Det er et slags Jevons paradoks (mer om dette i kapittel 6.5)

Hvorfor er det relevant?

Peak cycle er toppen i en konjunktur eller sektorsyklus. Økonomisk er det det punktet der GDP, produksjon og sysselsetting når platå før nedgang. Industrielt i minnebransjen (DRAM/HBM) betyr dette rekordhøye priser, utsolgte fabrikker, høye marginer. Alt før ny kapasitet kommer og skaper overkapasitet.

I minne-sektoren har dette vært normen i 3–5 års bølger. Men endrer reglene seg, og er verdens dyreste setning: «denne gangen er annerledes», relevant?



Kilde: Investopedia - Understanding Business Cycles: Phases and Measurement

Gavin Baker, CIO i Atreides, påpeker at dagens setup i minnemarkedet med høye priser, sterkt sentiment og økende tilbud historisk sett har vært et klassisk signal om å selge. Dette mønsteret har gjentatt seg i nesten alle minnesykluser de siste 25 årene. Unntaket er midten av 1990-tallet, den siste reelle kapasitetssyklusen, der etterspørselen eksploderte strukturelt fordi internettet krevde helt ny datainfrastruktur fra grunnen av.

Baker mener AI kan representere en tilsvarende unik syklus, ikke en vanlig boom-bust som vi har snakket om, men en utbygging av kapasitet der etterspørselen er strukturell. HBM skiller seg for øvrig betraktelig fra tradisjonell DRAM ved å være

skreddersydd og co-designet med GPU-produsentene, noe som gjør det mer til en kritisk og kunde-spesifikk komponent enn en ellers ren vare.

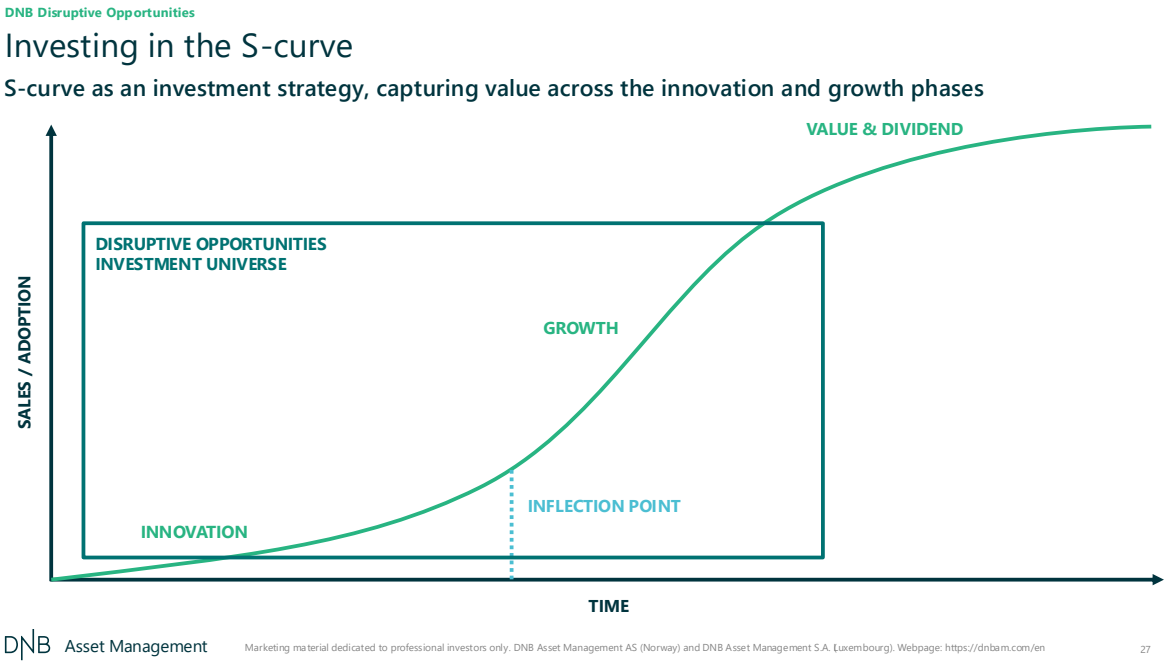
S-kurven som alternativ vekstmodell

Og mens peak cycle er en repetitiv bølge, er S-kurven en enveis adopsjonskurve for teknologi. Dette er noe vi har skrevet og snakket på inn og utpust i løpet av de siste 6 årene, så vi holder det kort denne gangen. Tre faser:

1. Langsom start (early adoption) - lav vekst, usikkerhet.

2. Rask vekstfase - inflection point, rask skalering.

3. Platå (maturity) - metning, eller ny S-kurve via innovasjon.

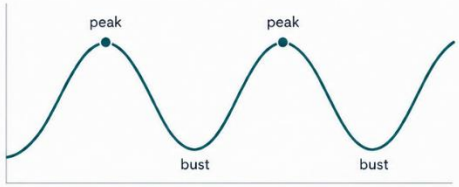


HBM representerer en ny S-kurve. Etterspørselen er strukturell og pris-uelastisk, ikke bare syklisk.

KLASSISK MØNSTER

Peak Cycle

Repetitiv bølge. Toppen er et absolutt maksimum før en kraftig nedtur.



Tid – 2 til 5 år per syklus

● Kurs

AI-SUPERSYKLUS

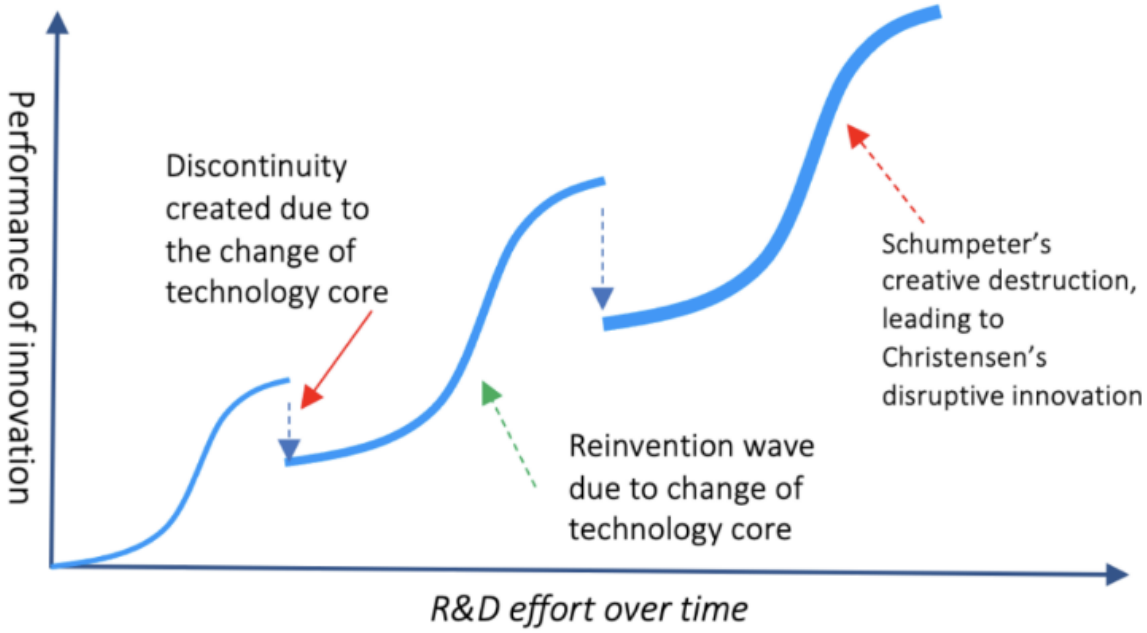
S-kurve

Én retning, i bølger. Første vekstfase flater ut med en mild dip, før en ny S-kurve tar den videre til et enda høyere nivå.

Tid – 5 til 20+ år

● Adopsjon / verdiskaping

ASPEKT	PEAK CYCLE (KLASSISK)	S-KURVE (AI-SUPERSYKLUS)
Form	Repetitiv bølge	Én retning — langsom → bratt → platå
Tidsramme	2–5 år per syklus	5–20+ år
Ved «peak»	Absolutt topp før bust	Inflection point (bratteste vekst)
Etter «peak»	Kraftig nedtur	Mild dip, men på høyere nivå



6.3 Syklusen prøver å temme seg selv

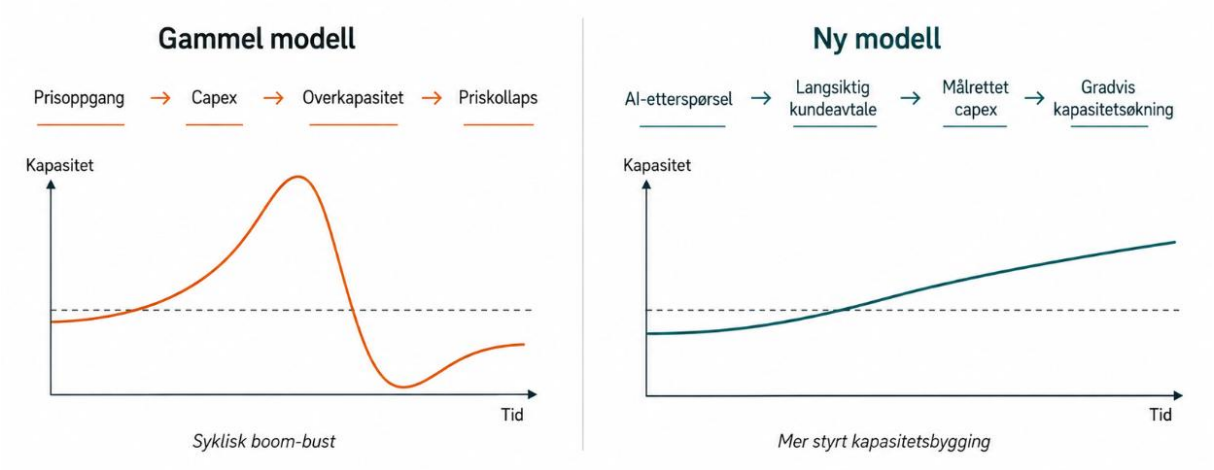
Et klassisk paradoks i minneindustrien er at gode tider ofte skaper sin egen undergang. Høye priser gir høy capex. Høy capex gir høy kapasitet. Høy kapasitet gir lavere priser.

Men denne gangen skjer noe nytt mellom første og andre ledd. Nemling at kundene binder seg.

Microns langsiktige avtaler dekker etter selskapets egne utsagn en stadig større del av virksomheten, med bindende kjøpsforpliktelser som i mange tilfeller strekker seg til 2030. Sandisk gjør noe lignende i NAND. På investordagen i august fortalte selskapet at åtte kunder allerede har inngått deres nye flerårige forretningsavtaler. Disse avtalene ventes å dekke rundt halvparten av bit-kapasiteten i FY2027 og rundt to tredjedeler i FY2028.

Det er interessant fordi kontraktene gjør noe produsentene historisk har manglet. De gir visibilitet.

Hvis du vet mer om etterspørselen fire år frem i tid, trenger du ikke bygge en fabrikk basert på håp om hvor markedet befinner seg når den åpner. Kapasiteten kan i større grad knyttes til faktisk kundebehov. Det kan bety at svingningene blir mindre.



For en industri bygget rundt boom og bust er den forskjellen ganske enorm.

Høy capex og lange kontrakter

De tre store (Samsung, SK Hynix, Micron) investerer massivt. Micron hever capex til over 25 milliarder dollar og Samsung planlegger selv 73 milliarder. Dette støtter peak cycle-tesen. Historisk har slik aggressivitet ført til overkapasitet. Kanskje i 2027–2028.

Men, og det er et stort men. Forskjellen er at det skjer et skift til lange kontrakter. Micron har signert sine første femårige avtaler og solgt ut hele 2026-produksjonen. Samsungs co-CEO Jun Young-hyun sier: «*Shifting from quarterly to 3–5-year contracts… to even out the peaks and troughs*». SK Group styreleder Chey Tae-won advarer om *wafer-shortage* på over 20% til 2030. Micron rapporterer at de nå kun klarer å møte 50% av faktisk etterspørsel.

Capex vs kontraktslengde

	Korte kontrakter	Lange kontrakter
Høy capex	**Klassisk peak + kraftig burst** Høye investeringer uten kontraktvern gir full eksponering mot etterspørselssjokk når syklusen snur	**Stabilisert sypersyklus** Investeringene er høye, men kontraktene låser etterspørselen over flere år og demper svingningene
Disiplinert capex	**Vedvarende undersupply** Lav investeringsvilje og korte avtaler gjør at tilbudet henger etter etterspørselen over tid	**Langvarig boom + prisstyrke** Begrenset tilbudsvekst kombinert med langsiktige avtaler gir varig prisdisiplin og marginstyrke

Hvorfor faller flere av aksjene på sterke tall? Markedet tror på syklus, men realiteten kan heller bli en ny supersyklus S-kurve.

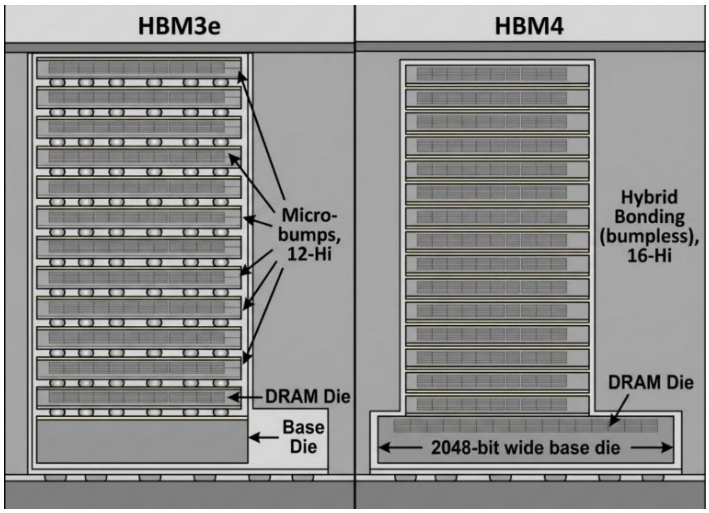
6.4 De teknologiske enablerne

Memory Wall oppstår fordi prosessorene har blitt langt raskere enn minnet klarer å levere data. Løsningen handler derfor om å gjøre veien mellom minne og prosessor bredere, kortere og mer effektiv.

1. HBM4 gir prosessoren mer data

HBM4 dobler interfacet til 2048 bit. Det betyr at langt mer data kan flyttes mellom minnet og prosessoren samtidig. Båndbredden øker til rundt 2 til 3,3 TB/s per stack.

Tenk på det som å gjøre veien inn til prosessoren bredere. Flere kjørefelt gjør at mer data kan passere samtidig, uten at prosessoren må vente like mye på minnet.



2. Flere lag på samme areal

HBM bygges ved å stable DRAM-lag oppå hverandre. HBM3E bruker typisk opptil 12 lag, mens HBM4 beveger seg mot 16-Hi.

Fordelen er at kapasiteten kan økes uten at minnet tar tilsvarende mer plass rundt prosessoren. Tenk skyskraper fremfor villaer. Samme tomt, men langt mer kapasitet i høyden.

3. TSV kobler lagene sammen

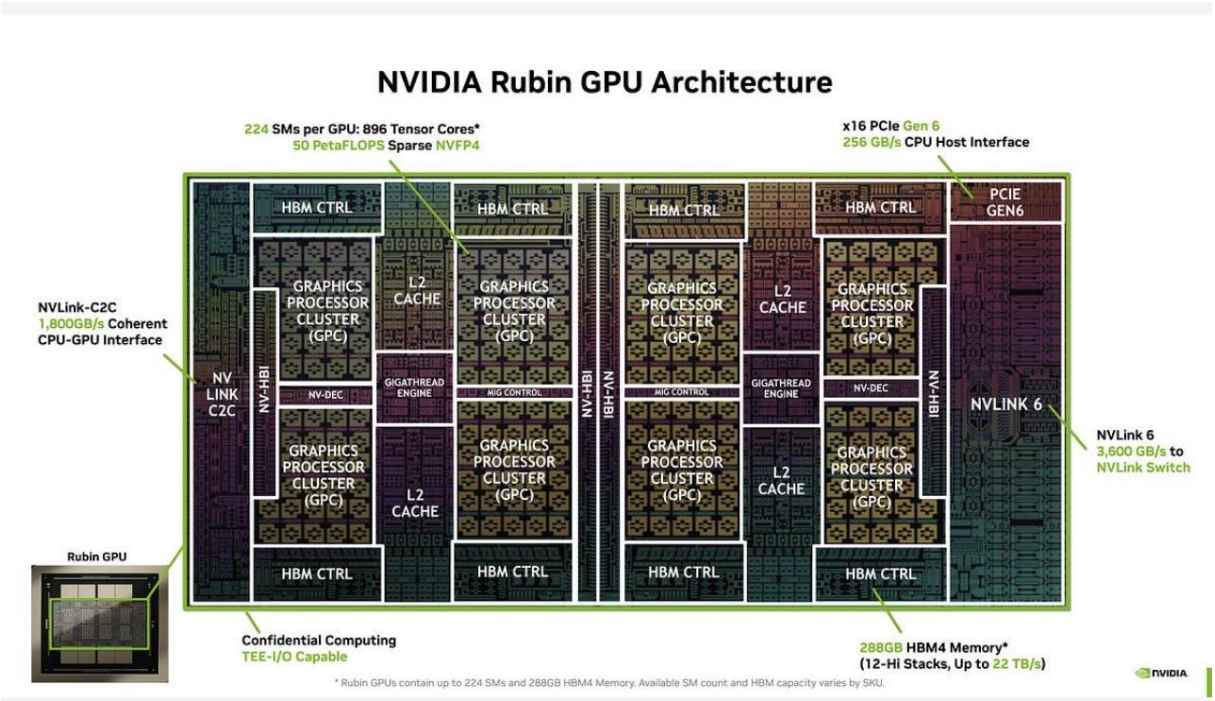
Når minnelagene stables, må de fortsatt kunne kommunisere raskt. TSV'er er små elektriske forbindelser som går vertikalt gjennom silisiumet og binder lagene sammen.

Det gjør at data kan bevege seg direkte gjennom stacken, i stedet for å måtte gå lengre veier rundt hver chip. Resultatet er høy dataflyt over svært korte avstander.

4. Hybrid bonding erstatter micro-bumps

I HBM3E kobles lagene sammen med små metallkontakter kalt micro-bumps. De fungerer godt, men tar plass mellom lagene og blir vanskeligere å skalere når stackene blir høyere.

HBM4 beveger seg mot hybrid bonding, der kobberflatene bindes langt tettere sammen uten de samme bumpene mellom hvert lag. Det gjør det mulig å redusere avstanden mellom lagene, bygge høyere stacks og forbedre forbindelsen gjennom hele minnepakken.



Konklusjonen er …

Vi mener ikke at vi er ved en klassisk peak, men midt i en hybrid sypersyklus. Lange kontrakter, wafer-bottleneck og HBM4-kompleksitet tipper balansen mot vedvarende vekst til 2030. Capex er høyt, men disiplinert og målrettet på high-value AI-minne.

Noen viktige tall å følge med på i tiden fremover er HBM4-yields i Q2-earnings, 2027-capex-guidance og Rubin-ramp. Minne er ikke lenger commodity, det er strategisk AI-infrastruktur. S-kurven vinner over den gamle syklusen. AI-revolusjonen trenger disse enablerne, fordi uten dem stopper alt. Med dem fortsetter boomen lenger og høyere enn noen klassisk modell spår.

6.5 AI-Supercycle møter Jevons' paradoks

I 2026 rapporterer Micron rekordsterk etterspørsel etter minne: Kundene får bare rundt 50 % av det de vil ha, Samsung og SK Hynix inngår 3-5 års kontrakter, og mangelen kan vare til 2030. Likevel falt aksjen på kvartalsrapporten. Investorer priser inn en klassisk peak cycle, at vi er på toppen av en syklus der boomen snart snur ned. Men er det riktig? Eller er vi midt i en strukturell AI-supercycle der nye teknologier som HBM4 og TurboQuant forandrer reglene? Vi oppsummerer hele dette hotte temaet Memory Wall som grunnproblem, via hardware- og software-løsninger, til Jevons' paradoks som viser hvorfor effektivitet ofte fører til mer, ikke mindre, etterspørsel og bruk.

Jevons' paradoks (utarbeidet i 1865) sier at når en ressurs blir mer effektiv, øker ofte totalforbruket. William Stanley Jevons så at mer effektive dampmaskiner førte til mer, ikke mindre, kullforbruk fordi industrien eksploderte.

Så hvis TurboQuant gjør inferens 8x billigere. Hva fører det til? Flere queries, lengre kontekst, flere modeller, edge-AI på telefoner og i roboter. Selv om hver query bruker 6x mindre minne, kan totalt HBM/DRAM-forbruk øke dramatisk fordi volumet vokser 10-100x.

Et billedlig eksempel kan være når en bil går fra 0,8 til 0,4 liter per mil. Da kjører folk ikke mindre, men mer, og tar lengre turer, kjøper større biler osv.. Totalt bensinforbruk stiger.

Kort oppsummert til slutt

Memory Wall er en av de største begrensningene i AI-utbyggingen. HBM4 angriper problemet fysisk. TurboQuant gjør det gjennom software. Når AI samtidig blir billigere å bruke, kan etterspørselen vokse raskere enn effektiviteten reduserer minnebehovet per oppgave.

Derfor mener vi dagens minnemarked bør forstås i lys av mer enn den klassiske boom-bust-syklusen. AI bygger en ny etterspørselsbase, og minne blir en stadig viktigere og mer strategisk del av selve infrastrukturen.

Og, jo mer autonome de fire rytterne blir, desto viktigere blir evnen til å holde kontekst over tid og intelligens i edgen.

Vi startet med menneskelig hukommelse og endte i silisium. Kanskje ligger noe av den neste store AI-utviklingen nettopp i koblingen mellom de to, si. Sakte, sakte, plutselig.

Og husk; Vi forvalter aksjer, ikke sannheter. Råd eller anbefalinger gis ikke, men vi deler våre perspektiver.

Takk for tiden.

Vi sees i fremtiden.

Mer gratis info?

lensikten med Disruptive Perspektiver:

år vi analyserer ulike temaer bruker vi mye tid og mange verktøy (kvartalsrapporter, analyser, dialog med eskapene, bedriftsbesøk, excel, kalkulator og ordmodeller). Ofte lager vi små notater, og noen ganger tore notater som vi tenker på som perspektiver. Vi er gamle nok til å vite at det sjeldent finnes sannheter, fte bare ulike perspektiver.

isruptive Perspektiver har kun én hensikt: Å dele våre perspektiver på temaer som former vår fremtid. ette er ikke akademiske notater, innlegg til et leksikon eller anbefalinger om å gjøre noe, kjøpe eller selge oe. Kun god gammeldags informasjonsdeling for å synliggjøre hvordan vi ser på ulike temaer på ubliseringstidspunktet. Perspektiver bli ikke mindre, kanskje heller mer verdt, når man deler det. Med det utgangspunktet; ha en fin reise i våre perspektiver.

Følg med gjennom LinkedIn, Substack og X

- Følg med på X, Substack og LinkedIn
- Skriver ofte om disrupsjoner og aksjemarkedet

....samt podkasten *Utbytte*

DNB Asset Management