

Lauri Hellsten

# NÄYTÖNOHJAINTEN EVOLUUTIO

Kandidaatintyö  
Informaatioteknologian ja viestinnän tiedekunta  
Tarkastaja: Katja Laine  
Joulukuu 2024

# TIIVISTELMÄ

Lauri Hellsten: Näytönohjainten evoluutio  
Kandidaatintyö  
Tampereen yliopisto  
Sähkötekniikan tutkinto-ohjelma  
Joulukuu 2024

---

Tämän työn tarkoituksena oli selvittää miten irralliset näytönohjaimet ovat kehittyneet, mistä ne ovat lähteneet liikkeelle ja millaiset tulevaisuuden näkymät niillä on. Työ on luonteeltaan kirjallisuuskatsaus ja lähteitä on etsitty valmistajien omilta sivuilta, tieteellisistä tietokannoista, uutisista sekä muista lähteistä kuten Wikipediasta, mikäli historiallista tietoa ei ole muualta ollut kunnolla saatavissa.

Näytönohjaimet ovat integroituja-, erillisiä samalle piirilevyille juotettuja- tai kokonaan irrallisia erillisiä korttimallisia komponentteja, joiden vastuulla on grafiikan tuominen näyttöpäätteelle. Näytönohjaimia löytyy nykypäivänä lähes jokaisesta laitteesta, esimerkiksi puhelimesta, tietokoneista ja pelikonsoleista. Näytönohjaimet eivät kuitenkaan ole aina olleet samanlaisia. Ensimmäiset näytönohjaimet olivat lähinnä kuvapuskureita, jotka saivat tulostettua vain rajatun määrän merkkejä. Ajan kuluessa näytönohjaimet alkoivat kykenemään tulostamaan jo kohtuullisen tehokkaasti liikkuva kuvaa yksittäisten merkkien sijaan. Ensimmäinen virallinen näytönohjain oli Nvidian GeForce 256 vuonna 1999. Tämän jälkeen näytönohjainten kehitys alkoi kiihtymään vuosi vuodelta, ja nykyään ollaan jo pisteessä, jossa näytönohjaimet suorittavat paljon muitakin tehtäviä kuin grafiikan piirtämistä. Tällaisia tehtäviä ovat esimerkiksi tekoälylaskenta, koneoppiminen, suurien datamäärien käsittely sekä tekoälyn integrointi erilaisiin puolustusteollisuuden järjestelmiin. Näytönohjaimien tulevaisuus on puhtaasti arvailun varassa, mutta lähitulevaisuutta voidaan kuitenkin arvioida jo nykypäivän tarpeiden ja kehitteillä olevien teknologioiden avulla.

Näytönohjaimien merkitystä ei voida aliarvioida, ne ovat mahdollistaneet hyvin suuren osan viimevuosien teknologisista saavutuksista suuren laskentatehonsa ansiosta, vaikkei sitä aina tulekaan ajatelleeksi. On kuitenkin muistettava, että näytönohjaimet ovat vielä suhteellisen uusi keksintö suuremmissa mittakaavassa. Vain aika voi näyttää kuinka pitkälle näytönohjaimet voidaan viedä, ennen kuin niiden kehitys jää ajastaan jälkeen.

Avainsanat: GPU, näytönohjaimet, evoluutio

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

# TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessäni on käytetty tekoälysovelluksia:

- Ei
- Kyllä

Ilmoitukseni mukaan olen käyttänyt opinnäytteessäni tutkielmaprosessin aikana seuraavia tekoälysovelluksia:

ScopusAI

Käyttötarkoitus: Apuna relevanttien lähteiden etsinnässä Google Scholarin ja Google Haun rinnalla

Osiot, joissa tekoälyä on käytetty: Luvusta viisi (5) eteenpäin työn loppua kohti

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien osat, joissa on hyödynnetty tekoälyä, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

# SISÄLLYSLUETTELO

1. JOHDANTO .....	1
2. NÄYTÖNOHJAINTEEN ALKUTAIVAL .....	2
2.1 Ensimmäisten näytönohjainten kehitys (1980- ja -90-luvut) .....	2
2.2 2D-grafiikan tuottaminen ja aikakauden haasteet .....	2
2.3 Yksinkertaisten graafisten prosessorien toiminta .....	3
3. 3D-GRAFIIKAN VALLANKUMOUS .....	4
3.1 3D-grafiikan yleistymisen 1990-luvun lopulla .....	4
3.2 GPU:n (Graphics Processing Unit) synty ja ensimmäiset mallit .....	5
3.3 Edellytykset teknologian kehittymiselle .....	6
4. MARKKINOIDEN LYHYT HISTORIA .....	7
4.1 Keskeiset toimijat historiassa .....	7
4.2 Suurten kilpailijoiden innovaatioita .....	8
4.3 Tärkeitä malleja markkinoilla .....	9
5. NÄYTÖNOHJAINTEEN TEKNINEN KEHITYS .....	11
5.1 GPU-arkkitehtuurin kehitys .....	11
5.2 Ydinten ja laskentatehon lisääntyminen .....	13
5.3 Muistin ja nopeuksien kasvaminen .....	15
5.4 Lämmönhallinta ja virrankulutus .....	17
5.5 Liitännät .....	19
6. NÄYTÖNOHJAINTEEN KÄYTTÖ PELAAMISEN ULKOPUOLELLA .....	22
6.1 GPU datatieteessä, tekoälyssä ja laskennallisessa tieteessä .....	22
6.2 GPU:n rooli lohkoketjujen louhinnassa .....	23
6.3 Näytönohjainten merkitys ja käyttö uusilla aloilla .....	24
7. NYKYISET JA TULEVAISUUDEN TRENDIT .....	25
7.1 Ray Tracing ja muut teknologiat .....	25
7.2 Pilvipohjainen grafiikan prosessointi .....	26
7.3 Tulevaisuuden näköalat: kvanttilaskenta ja muut mahdollisuudet .....	27
8. YHTEENVETO .....	29
9. LÄHTEET .....	30

## LYHENTEET JA MERKINNÄT

2D	Kaksiulotteinen grafiikka
3D	Kolmiulotteinen grafiikka
AFR	Alternate Frame Rendering
AGP	Accelerated Graphics Port
API	Application programming interface, Ohjelmointirajapinta
CGA	Color Graphics Adapter
CPU	Graphical Processing Unit, Keskusprosessori
DDR	Double Data Rate
DRAM	Dynamic Random-Access Memory
FPS	Frames Per Second
GDDR	Graphics Double Data Rate
GPGPU	General Purpose Graphics Processing Unit
GPU	Graphical Processing Unit, Näytönohjain
HGC	Hercules Graphics Controller
I/O	Input/Output
IoT	Internet of Things
MDA	Monochrome Display Adapter
MT/s	Millions of Transfers per second
PCIe	Peripheral Component Interconnect Express
PoW	Proof of Work
RAM	Random-Access Memory
RegF	Register File
ROM	Read-only memory
SDRAM	Synchronous Dynamic Random-Access Memory
SFR	Split Frame Rendering
SM	Streaming Multiprocessor (Maxwell arkkitehtuuri ja eteenpäin)
SMX	Streaming Multiprocessor (Kepler arkkitehtuuri)
TTL	Transistor-transistor logic

# 1. JOHDANTO

Nykymaailmassa on erittäin todennäköistä, että taskuistasi tai kotoasi löytyy laite, joka pystyy näyttämään asioita elottomalla näyttöpaneelilla ilman liikkuvia osia ja laskemaan todella suuria määriä dataa lyhyessä ajassa. Harva meistä tulee kuitenkaan aktiivisesti ajatelleeksi, kuinka tähän pisteeseen on päädytty ajasta, jolloin Tennis for Two-nimellä tunnettu simulaatio tenniksestä pyöri analogisen tietokoneen voimin oskilloskoopin näytöllä [1]. Nykypäivänä tällainen tapa visualisoida esimerkiksi videopelejä tuntuu absurdilta, sillä näytönohjaimet suorittavat käytännössä kaiken graafisen laskennan hyvin suurilla tarkkuuksilla jo puhelimen kokoluokassa.

Näytönohjaimet (GPU, Graphics Processing Unit) ovat rinnakkaiseen laskentaan perustuvia mikroprosessoreja. Niiden tarkoituksena on siirtää 2D- ja 3D-grafiikan renderöinti, eli tuottaminen, pois keskusprosessorilta (CPU, Central Processing Unit) ja nopeuttaa täten grafiikan tuottamista [2]. Arkikielessä näytönohjaimella tarkoitetaan usein mikroprosessorin sijaan koko näytönohjainpakettia. Näytönohjainta voidaan verrata isoon rahtilaivaan, joka kykenee siirtämään (prosessoimaan) paljon tietoa kerralla mutta ei välttämättä ole hirveän nopea. Keskusprosessori voidaan nähdä nopeana lentokoneena, joka kykenee siirtämään (prosessoimaan) vähemmän asioita samanaikaisesti mutta on paljon nopeampi ja joustavampi prosessien suhteen.

Koska lähes kaikki laitteet käyttävät jonkintasoista näytönohjainta, on niitä valmistettu useassa eri kokoluokassa ja useaan eri käyttötarkoitukseen. Puhelimeissa ja useissa tietokonesuorittimissa näytönohjain on integroituna suorittimeen tilan ja tehonkulutuksen minimoimiseksi. Tehokkaammissa pelitietokoneissa ja pöytäkoneissa näytönohjain on usein erillinen tiedonsiirtoväylään liitettävä tehokkaampi korttimallinen näytönohjain ja pelikonsoleissa näytönohjain on juotettuna samalle emolevyille kuin prosessori tilan säästämiseksi. Tunnetuimpia näytönohjaimia ovat nykypäivänä Nvidian GTX- ja RTX-sarja, AMD:n Radeon RX-sarja sekä Intelin Graphics- ja ARC-sarja.

Tässä kirjallisuuskatsauksessa tutkitaan miten erityisesti pöytätietokoneissa käytetyt ulkoiset näytönohjaimet ovat kehittyneet ajan saatossa, mistä ne ovat lähteneet liikkeelle ja miltä niiden tulevaisuus näyttää. Luvuissa 2,3 ja 5 käsitellään näytönohjainten teknologista kehitystä tähän päivään asti, luvussa 4 keskitytään kilpailuun näytönohjainmarkkinoilla, luvussa 6 tarkastellaan näytönohjaimien käytöstä pelaamisen sekä kuvan näytölle tuottamisen ulkopuolella ja lopuksi luvussa 7 nykyisiä ja tulevaisuuden trendejä.

## 2. NÄYTÖNOHJAINTEEN ALKUTAIVAL

Näyttöohjaimet, tai niiden ensimmäiset alkeelliset vastikkeet, eivät aina ole toimineet samalla periaatteella kuin ne toimivat nykypäivänä. Tässä luvussa selvitetään, millaisia ensimmäiset ulkoiset kuvan piirtämiseen tarkoitetut näyttöadapterit olivat.

### 2.1 Ensimmäisten näyttöohjainten kehitys (1980- ja -90-luvut)

Ensimmäiset näyttöohjainten vastikkeet näkivät päivänvaloa jo 1980-luvun alkupuolella. Nämä alkukantaiset näyttöohjaimet olivat lähinnä integroitua kuvapuskureja tai näyttöadaptereita, jotka olivat TTL-logiikkasiruista koostuvia tietokoneeseen liitettäviä kortteja. Nämä olivat lähinnä jatkeita prosessorille, joista osa kykeni piirtämään kuvia pelkästään ennalta määrättyjen symbolien lisäksi, esimerkiksi HGC (Hercules Graphics Controller) sekä CGA (Color Graphics Adapter) [2][6][7]. Hiljalleen teknologian kehittyessä alettiin siirtymään pois TTL-siruista koostuvista korteista kohti mikroprosessoreja.

Myöhemmin vuonna 1989 Silicon Graphics Inc. julkaisi nykyään käytetyimmän ja tuetuimman ohjelmointirajapinnan (API) OpenGL:n. Tämä API muovasi pitkälti tulevien näyttöohjaimien kehitystä ja samalla Silicon Graphics muotoili uudestaan niin sanottua grafiikkaliukuhihnaa [2].

1990-luvulla grafiikkaliukuhihna alkoi ottaa muotoaan ja tämän myötä näyttöadapterit alkoivat lähestymään nykyistä näyttöohjainten rakennetta. 1990-luvun puolivälin aikoihin markkinoilta alkoi löytyä jo kuluttajille helpommin saatavilla olevia näyttöadaptereita, kuten tunnetuimmat Nvidian Riva TNT, 3DFX:n Voodoo sekä ATI:n, nykyinen AMD, Rage. Nämä näyttöadapterit kykenivät kuitenkin tuottamaan vain yhden pikselin yhtä kellosykliä kohden, mikä loppujen lopuksi johti rinnakkaisten grafiikkaliukuhihnojen ja useampien ytimien lisäämiseen näyttöohjaimille. Tämä mahdollisti useampien pikselien tuottaminen yhtä kellosykliä kohden [2]. Vasta vuonna 1999 Nvidia julkaisi GeForce 256 näyttöohjaimen, joka virallisesti vakiinnutti termin näyttöohjain eli GPU:n [4].

### 2.2 2D-grafiikan tuottaminen ja aikakauden haasteet

1980-luvun alussa näyttöadapterien alkukantainen rakenne mahdollisti kortin mukaan joko yksinkertaisten symbolien tulostamisen tai nykystandardin mukaan hyvin alkukantaisia ja epätarkkoja kuvia näyttöadapterin suorittaessa ainoastaan kuvan renderöinnin,

jonka prosessori oli tuottanut. Vasta noin vuonna 1997 näytönohjaimet alkoivat suorittamaan muitakin kuvan tuottamiseen liittyviä tehtäviä pelkästään kuvan näytölle piirtämisen ohella [2].

Haasteita teetti mm. hidas kuvan tuottaminen, josta edellisessä luvun 2.1 lopussa mainittiin. Grafiikkaliukuhihna oli teknologian alkukantaisuuden takia hyvin kapea ja pikselitulosteisessa grafiikassa yhtä kello sykliä kohden saatiin tuotettua vain yksi pikseli [2]. Tämä tieto koskee saatavilla olevan tiedon perusteella vain 1990-luvun näyttöadaptereita kuten Riva TNT:ä. Esimerkiksi TTL-tyylisestä CGA-kortista ei löydy tarpeeksi tarkkaa tietoa, kuinka monta sykliä pikselin piirtämiseen kului.

Myös värit olivat erittäin rajatut aivan alkuaikoina ennen siirtymistä mikroprosessoripohjaiseen näytönohjainteknologiaan. CGA kykeni näyttämään teoriassa 16 eri väriä, mutta kaikki värit eivät olleet käytettävissä kaikilla resoluutioilla rajatun muistimäärän ja tiedon siirtolevyyden takia. Esimerkiksi resoluutiolla 160x100 käytettävissä oli koko 16 värin paletti, resoluutiolla 320x200 käytössä oli 4 väriä ja korkean resoluution 640x200 tilassa käytettävissä oli ainoastaan kaksi väriä, jotka voitiin valita vapaasti väripaletilta [7].

### **2.3 Yksinkertaisten graafisten prosessorien toiminta**

Esimerkkinä yhtenä ensimmäisistä näyttöadaptereista voidaan pitää IBM:n MDA-korttia (Monochrome Display Adapter), joka pystyi tuottamaan vain ROM-muistiin ennalta määrittäjä symboleja kahdella eri intensiteetillä, muuttamaan symboleista alleviivattuja tai tulostamaan ne paksumpina. MDA koostui useista logiikkapiireistä, 4 kilotavusta RAM-muistia, ROM-muistista merkeille ja 6845 CRT-ohjaimesta [8]. Toinen IBM:n samoihin aikoihin julkaisema grafiikkakortti oli jo mainittu CGA, joka kykeni näyttämään myös pikselikuvia. Tämä oli mahdollista, sillä kortti kykeni muodostamaan bitmap-tiedostoja, joka käytännössä vastasi näytöllä näkyvää kuvaa, jolloin jokainen yksittäinen pikseli oli mahdollista määrittää erikseen. CGA perustui myös 6845 CRT-ohjaimen [7]. Molempien korttien kohdalla ainut keino muuttaa merkkisarjaa oli vaihtaa ROM-muisti. Mutta koska CGA pystyi piirtämään yksittäisiä pikseleitä, oli käytännössä mahdollista piirtää näytölle mitä tahansa symboleja. ROM-muisti olikin lähinnä tekstiä varten, joka saatettiin valita CGA-kortilla kuvan piirtämistilan sijaan.

### 3. 3D-GRAFIIKAN VALLANKUMOUS

1990-luku voidaan nähdä aikakautena, jolloin esimerkiksi internet ja viihdemedia lähtivät todelliseen nousukiitoon. Internet avautui yleiseen käyttöön 1.1.1993 kun sitä edeltänyt ARPANET ja Defense Data Network siirtyivät käyttämään TCP/IP (Transfer Control Protocol/Internetwork Protocol) protokollaa, joka mahdollisti kaikkien erilaisten tietokoneiden välisen viestinnän verkon ylitse [15]. Televisiosarjat yleistyivät, animoituja elokuvia, kuten Toy Story [9], ilmestyi vuonna 1995 elokuvateattereihin ensimmäisenä puhtaasti tietokoneella animoituina täysipitkinä 3D-elokuvana. Sekä viihdemedian että internetin avautuminen takasivat jälkikäteen katsottuna varman ja tukevan pohjan 3D-grafiikan sekä siihen tarvittavan teknologian kehittymiselle.

#### 3.1 3D-grafiikan yleistyminen 1990-luvun lopulla

Näytönohjainten kehittyessä 1990-luvulla, kehittyi samalla luonnollisesti grafiikka. Erityisesti videopelejä alkoi ilmestymään alati kiihtyvään tahtiin. Näitä olivat esimerkiksi Doom (1993) [10], Warcraft: Orcs and Humans (1994), Tomb Raider (1996), Legend of Zelda: Ocarina of time (1998) sekä Everquest (1999) [11]. Graafinen kehitys alkoi ottamaan jättimäisiä harppauksia, joka voidaan huomata vertailtaessa kuvia 1 ja 2, jotka ovat kuvankaappauksia 90-luvun alun Doomista ja 90-luvun lopun Zeldasta.



*Kuva 1. Kuvankaappaus 1993 vuoden Doom-peleistä [10].*



*Kuva 2. Kuvankaappaus 1998 vuoden The Legend of Zelda-pelistä [12].*

### 3.2 GPU:n (Graphics Processing Unit) synty ja ensimmäiset mallit

GPU-termin kehitti Sony jo vuonna 1994 ensimmäisen Playstation-konsolinsa yhteydessä [3]. Termi GPU kuitenkin saavutti nykyisen statuksensa vasta kun Nvidia julkaisi vuonna 1999 ensimmäisen ulkoisen korttimuotoisen näytönohjaimensa, joka sai nimen GeForce 256. Tämä näytönohjain suoritti koko grafiikkaliuokuhinnan yhdellä sirulla, joka liitettiin emolevyllä löytyneeseen AGP 4X väylään. AGP väylän korvasi myöhemmin PCIe väylästandardi. [4] Kuvassa 4 on esiteltyä Nvidian GeForce 256.



*Kuva 3. Nvidia GeForce 256, koodinimeltään NV10 [5].*

GeForce 256 sisälsi 256-bittisen 3D prosessorin, 32 megatavua SDRAM muistia (Synchronous Dynamic Random-Access Memory), pystyi tuottamaan 4 pikseliä per kello sykli ja ennen kaikkea suorittamaan renderöinnin, geometriamuunnokset, valaistukset sekä rasteröinnin kaiken yhdessä paketissa [14].

Vuonna 2001 Nvidia julkaisi GeForce 3 näytönohjaimen ja samaan aikaan toinen suosittu näytönohjain oli ATI:n Radeon 8500. Seuraavana vuonna Nvidia julkaisi GeForce FX näytönohjaimensa ja ATI puolestaan Radeon 9700:n.

### 3.3 Edellytykset teknologian kehittymiselle

On oleellista ymmärtää, että näytönohjaimien teknologisista harppauksista puhuttaessa on huomioitava myös näytönohjainta ympäröivien ja tukevien komponenttien kehitys.

1990-luvun alkupuoliskolla markkinoilla koettiin suuri harppaus, kun huomattava määrä eri mikrotietokoneita, eli henkilökohtaisia tietokoneita, jotka nykypäivänä vastaavat tavallisia tietokoneita [16], ilmestyi markkinoille erilaisilla ominaisuuksilla eri ostajakunnille. Tietokoneista tuli tehokkaampia, ominaisuuksiltaan rikkaampia ja niitä oli saatavilla modulaarisessa muodossa hinnan samalla laskiessa. [17, s.11] Tätä on jatkunut pitkälti samalla kaavalla nykypäivään asti, vaikka viime aikoina on nähty myös välillä hintojen nousua johtuen osittain hetkellisestä mikrosirujen puutteesta markkinoilla 2020-luvun alussa sekä markkinoiden lähes monopolisoitumisesta keskusprosessorien ja näytönohjaimien puolella. Yleisesti hintojen lasku ja ominaisuuksien lisääntyminen on tuonut teknologian yhä useampien saataville, joka on entisestään mahdollistanut yhä useampien tahojen pääsyn kehittämään uutta teknologiaa.

Kuten mainittu, näytönohjaimien kehittyminen vaatii oleellisesti muidenkin komponenttien kehittymistä samalla sekä näytönohjainkortilla kuin sen ulkopuolellakin. Tämä koskee oleellisimmin parempaa tehonvälitystä näytönohjaimelle ja muille komponenteille, muistien nopeuksien ja kapasiteettien kasvua, keskusprosessorien nopeuksien nousua sekä komponenttien välisen tiedonsiirron kehittymistä.

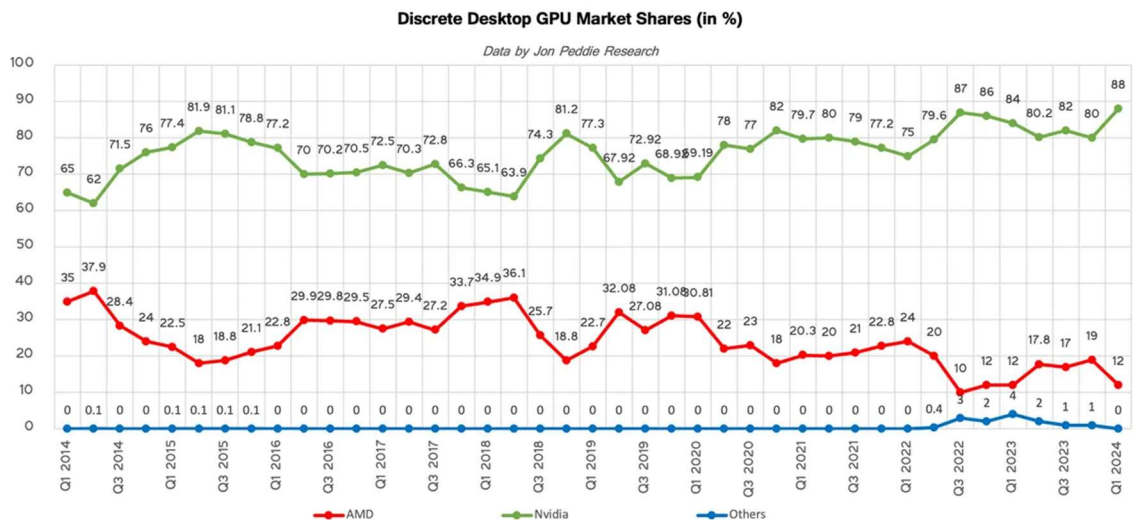
Ilman muiden komponenttien kehitystä saavutettaisiin jossain vaiheessa piste, missä näytönohjaimista ei voitaisi saada enempää hyötyä irti. Tätä kutsutaan pullonkaulaksi, englanniksi *bottleneck*. Yksi tunnetuimmista pullonkauloista erityisesti pelitietokoneita rakentavien keskuudessa on keskusprosessorin ja näytönohjaimen välinen pullonkaula. Alitehoinen keskusprosessori yhdistettynä ylitehokkaaseen näytönohjaimeen aiheuttaa pullonkaulan, josta käytetään usein termiä *CPU bottleneck*. Sama toimii myös toisinpäin näytönohjaimen ollessa alimitoitettu, jolloin kyseessä on *GPU bottleneck*.

## 4. MARKKINOIDEN LYHYT HISTORIA

Tähän mennessä on lähinnä puhuttu yksittäisistä malleista, miten näytönohjaimet ovat kehittyneet alkukantaisista kuvapuskureista lähemmäs nykyisiä näytönohjaimia Nvidia GeForce 256 näytönohjaimeen asti. Tässä kappaleessa kerrotaan näytönohjainten historiasta markkinoiden ja valmistajien näkökulmasta. Huomioitavaa on, että valmistajista puhuttaessa tarkoitetaan näytönohjaimen prosessorin valmistajaa, eikä korttien valmistajia, jotka ostavat sirut, joista sitten nämä tekevät lopullisen näytönohjaimen.

### 4.1 Keskeiset toimijat historiassa

Nykypäivänä markkinoilta voidaan sanoa löytyvän näytönohjaimia pelkästään kolmelta valmistajalta, joista Nvidialla oli hallussaan 88 % markkinaosuudesta vuoden 2024 alussa, AMD:lla noin 12 % ja Intelin osuus on ”häviävän pieni” [18]. Intel liittyi kuitenkin vasta muutama vuosi sitten kilpailuun ARC sarjallaan, joka ei ole vielä saanut samanaista lojaalia käyttäjäkuntaa kuten Nvidia tai AMD. Kuvassa 5 on esiteltynä erillisten näytönohjainten prosentuaaliset markkinaosuudet kymmenen vuoden ajalta väliiltä 2014–2024.



**Kuva 4.** Nvidian, AMD:n sekä muiden markkinaosuudet viimeisen kymmenen vuoden aikana. Intel on sisällytettyä muut kategorian alle. [18]

Oleellista on huomioida kuvaa tulkittaessa, että siinä puhutaan erillisistä näytönohjaimista, eikä esimerkiksi keskusprosessoriin sisäänrakennetuista, eli integroiduista, näytönohjaimista. Mikäli nämä otettaisiin huomioon, kasvaisi sekä AMD:n että Intelin osuudet merkittävästi, sillä todella isossa osassa kotikäyttöön tarkoitetuissa kannettavissa tietokoneissa on integroitu näytönohjain erillisen sijaan. Koska kolmesta edellä mainitusta valmistajasta vain Intel ja AMD valmistavat myös keskusprosessoreja, pääsevät nämä sisällyttämään näytönohjaimiaan laitteisiin, joissa ei ole käyttötarvetta ulkoisille ohjaimille, toisin kuin Nvidia.

Ajassa taaksepäin mennessä valmistajia on kuitenkin ollut paljon enemmän kuin tähän mennessä on mainittu, joista tunnetuimpia ovat olleet muun muassa 3dfx, ATI sekä Real3D kymmenien muiden valmistajien joukossa. Näistä kuitenkin käytännössä kaikki katosivat markkinoilta ajan kuluessa esimerkiksi jouduttuaan ostetuksi ja liitetyksi osaksi ostajaansa tai jouduttuaan konkurssiin. Näistä esimerkiksi 3dfx menetti teknologiansa Nvidialle jouduttuaan konkurssiin, Real3D:n osti Intel ja ATI:sta tuli nykyinen AMD. [19] Ainoastaan Nvidia sekä AMD ovat selvinneet aloitettuaan, vaikka AMD tunnettiin aluksi eri nimellä.

## 4.2 Suurten kilpailijoiden innovaatioita

Nvidia on vuosiansa aikana ollut iso innovaattori alallaan. Osa innovaatioista on luonnollisesti ollut mahdollista vasta ostettuaan kilpailijoitaan, mutta lasketaan ne silti Nvidian saavutusten kasaan. Ensimmäisenä ja todennäköisesti merkittävimpana innovaationa Nvidialta voidaan pitää GeForce 256 näytönohjainta, joka vakiinnutti GPU-termin ja oli iso teknologinen harppaus 3D-grafiikassa. Myöhemmin vuonna 2006 Nvidia esitteli maailmalle osaksi näytönohjainsiruja liitetyt CUDA-ytimet, jotka mahdollistivat rinnakkaislaskennan ennennäkemättömillä nopeuksilla. Tämä nopeutti erityisesti tieteellistä laskentaa ja simulointia sekä videon käsittelyä mullistaen samalla käsityksen näytönohjainten mahdollisuuksista. Vuonna 2012 Nvidia julkaisi Kepler-arkkitehtuurin, parantaen jälleen laskentatehoa sekä virrankulutusta merkittävästi. Titan X-näytönohjain puolestaan vuonna 2014 oli ensimmäinen laatuaan sen tuodessa yli petaflopin laskentatehon, eli  $1000 * 10^{12}$  laskutoimitusta sekunnissa, kaikkien saataville. Tekoälyn puolella vuonna 2017 Nvidia julkaisi Volta-arkkitehtuurin ja tämän mukana tekoälylaskentaa kiihdyttämisessä loistavat Tensor-ytimet. Seuraavana vuonna esitelty Turing-arkkitehtuuri puolestaan vei Tensor-ytimien suorituskykyä pidemmälle, kun ne optimoitiin syväoppimiseen liittyviin

tehtäviin. [s.7, 20] Lisäksi yksi viimeisimmistä merkittävistä virstanpylväistä oli ensimmäisen RTX-näytönohjainsarjan julkaisu, joka mahdollisti reaaliaikaisen säteenseurannan, eli Ray Tracingin.

AMD:lla puolestaan merkittävimpiä innovaatioita ovat erityisesti vuonna 2021 julkaistu RDNA (Radeon DNA) arkkitehtuuri, jonka vuonna 2020 julkaisemaa versiota RDNA 2 hyödynnetään erityisesti Playstation 5- ja Xbox Series X-pelikonsoleissa [21]. Samana vuonna julkaistujen RX6000 sarjan näytönohjaimien ja RDNA 2 arkkitehtuurin mukana AMD julkaisi Infinity Cache muistin, joka lupasi erittäin korkeita kaistanleveyksiä ja pieniä vasteaikoja. Kauemmas historiaan mennessä muistettavaa on, että AMD:n näytönohjaimista puhuttaessa oli valmistaja vielä ATI, mutta yksinkertaisuuden puolesta ATI laskeetaan nyt AMD:ksi. Vuonna 2000 ATI julkaisi ensimmäisen Radeon brändätyn näytönohjaimensa, R100:n, joka oli varteenotettava kilpailija Nvidian GeForce 2:lle sekä 3dfx:n Voodoo5:lle. R100 oli aikanaan yksi teknologisesti kehittyneimmistä tavallisille kuluttajille suunnatuista näytönohjaimista. [23]

Intelin ollessa vielä erittäin uusi tulokas modernien ulkoisten näytönohjaimien rintamalla, saatavilla olevan informaation määrä innovaatioista on luonnollisesti pieni. Lisäksi Intel joutui käyttämään paljon resursseja ajuriongelmiin korjaamiseen näytönohjaimien julkaisun jälkeen.

### 4.3 Tärkeitä malleja markkinoilla

Tärkeän määritelmä on laaja ja voi erota eri tahojen välillä hieman. Tässä työssä tärkeä malli tai sukupolvi on määritelty yksinkertaisuuden puolesta sellaiseksi, joka on ollut teknologisesti huomattavan kehittynyt, synnyttänyt uusia standardeja ja tavoitteita tai muuten ohjannut näytönohjaimien kehityksen suuntaa merkittävästi. Lisäksi nyt on rajattu tarkastelun kohteiksi ainoastaan Nvidia sekä AMD pitkän historian takia.

Luonnollisesti ensimmäisenä on mainittava jo useasti mainittu GeForce 256, joka sementoi GPU termin ja oli ensimmäinen näytönohjain, joka sisällytti jo mainitusti koko grafiikkaliukuhinnan yhteen siruun.

Hieman myöhemmin 2000-luvun alkupuoliskolla Nvidia julkaisi GeForce 6800 sarjan näytönohjaimensa, tuodessaan mukanaan SLI (Scalable Link Interface) liitäntämahdollisuuden, joka pohjautui 3dfx:n Scan-Line interleave teknologiaan. Tämä mahdollisti kahden näytönohjaimen yhteen liittämisen parantaen huomattavasti suorituskykyä. Lisäksi kyseinen näytönohjain oli ensimmäisten joukossa käyttäessään silloin uutta ja nopeaa GDDR3 muistia. [24]

Kaksi vuotta myöhemmin Nvidia julkaisi GeForce 8800 näyttöohjaimensa. GeForce 8800 kykeni päihittämään edellisen sukupolven GeForce 6800 näyttöohjaimen sekä yksilökonfiguraatiossa sekä sen käyttäessä kahden näyttöohjaimen SLI-konfiguraatiota. GeForce 8800 oli myös ensimmäinen näyttöohjain, joka käytti niin kutsuttua "Unified shader" arkkitehtuuria. Unified shader arkkitehtuuri mahdollisti kaikkien 3D-grafiikkaan liittyvien laskujen suorittamisen yhden paketin sisällä, joka vähensi eri prosesseille tarvittavien piirien määrää. Lisäksi se mahdollisti myös universaalien laskutavan kortin laskutehollisuuden, FLOPS (Floating point operations per second), eli karkeasti laskutoimitusten määrä sekunnissa [25]. [24]

Vuonna 2016 Nvidia julkaisi GTX 1080 näyttöohjaimen, joka pärjäsi paljon uudemmalle 2021 julkaistulle Nvidian RTX 3060 näyttöohjaimelle. Siitä jopa tuli eräänlainen verrokki, johon muita näyttöohjaimia verrattiin vielä paljon sen julkaisunkin jälkeen. [24]

GTX 10xx-sarjan seuraajana julkaistu RTX-sarjan näyttöohjaimet toivat kerta toisensa jälkeen ennennäkemättömiä loikkia laskutehoissa, kuten 70 % suuremman suoritusnopeuden RTX 3080 kohdalla verrattuna RTX 2080 näyttöohjaimeseen jopa 4K grafiikoilla. RTX sarja oli myös ensimmäinen, joka mahdollisti reaaliaikaisen säteenseurannan. [24]

Siirryttäessä AMD:n tarjontaan, yhtiön ensimmäisenä 3D-kiihdyttimenä ja kilpailijana Nvidialle toimi vielä sen aikaisen ATI:n vuonna 1999 julkaistu Rage 128. Vaikka Rage 128 ei pärjännytkaan kilpaileville näyttöohjaimille, loi se silti polkua AMD:n tuleville näyttöohjaimille ja uusille versoille Rage 128:sta. [26]

Vuonna 2000 AMD julkaisi Radeon DDR korttinsa, joka toi paljon uusia ominaisuuksia, vaikkei ollutkaan laskentatehojen kärjessä. Radeon DDR oli kuitenkin ensimmäisiä kortteja, jolla oli DirectX 8 tuki, pärjäsi hyvin 32-bitin väriavaruuden kanssa sekä sisälsi videon sisään- ja ulostuonnin. [26]

Vuonna 2002 AMD julkaisi Radeon 9700 näyttöohjaimensa, joka kykeni vihdoinkin päihittämään kilpailevan GeForce 4 Ti 4600 näyttöohjaimen 3D-suorituskyvyssä. Lisäksi se oli ensimmäinen DirectX 9:ää tukeva näyttöohjain markkinoilla. [26]

Paljon myöhemmin vuonna 2019 AMD julkaisi RX 5000 sarjan näyttöohjaimet, jotka toivat AMD:n Nvidian rinnalle tehovertailussa. RX sarjassa käytettyä teknologiaa hyödynnetään myös viimeisimmissä pelikonsoleissa Nvidian sijaan. [26]

## 5. NÄYTÖNOHJAINTEN TEKNINEN KEHITYS

Näytönohjaimien tekninen kehitys voidaan tiivistää yksinkertaisesti arkkitehtuurin eri versioihin. Karkeasti jokainen näytönohjainsukupolvi on käyttänyt yhtä ja samaa arkkitehtuuria, jolloin ominaisuudet ovat olleet pitkälti samanlaisia sukupolven mallien välillä. Näytönohjain saakin ominaisuutensa arkkitehtuurilta. Tässä luvussa tutkitaan nyt itse arkkitehtuurien kehitystä sekä miten arkkitehtuurin muodostavat tekijät ovat kehittyneet itsessään. Lisäksi tutkimme lähinnä Nvidian arkkitehtuuria sen ollessa kypsien markkinoilla olevista vaihtoehdoista, ellei toisin mainita [27].

### 5.1 GPU-arkkitehtuurin kehitys

Näytönohjaimien arkkitehtuuri on fundamentaalisesti erilainen keskusprosessorien arkkitehtuurista. Näytönohjat perustuvat datan rinnakkaiseen prosessoimiseen eli korkean datamäärän kerralla prosessoimiseen. Keskusprosessorit perustuvat puolestaan tehtävien rinnakkaiseen prosessoimiseen.

Alkuaikoina näytönohjaimien toiminnot olivat pitkälti lukittuja suorittamaan lähinnä 3D-grafiikan tuottamiseen liittyviä laskutoimituksia ja prosessi oli tehty pyörimään grafiikkaliukuhinnan ympärille. Liukuhinnan vaiheet olivat erillisiä piirejä prosessorilla. Uudet ohjelmointirajapinnat avasivat kuitenkin jotain ominaisuuksia ohjelmoijien käyttöön. Ohjelmoitavat prosessorit alkoivat kuitenkin myöhemmin ilmestymään markkinoille, kun kehittäjät halusivat luoda monimutkaisempaa grafiikkaa ja visuaalisia tehosteita. Yksi ensimmäisistä yrityksistä tuoda markkinoille ohjelmoitavampaa näytönohjainta oli GeForce 3 näytönohjain, joka mahdollisti esimerkiksi ohjelmoitavat vertex shaderit. [27] Vertex shader on lyhyesti funktio, joka mahdollisti grafiikassa käytettyjen kolmion muotoisten elementtien kulmien kärkipisteiden manipuloinnin. Manipulointi saattaa koskea esimerkiksi sijaintia, väriä tai valaistusdataa. [28] Jo aikaisemmin mainittu GeForce 8800 oli puolestaan ensimmäinen näytönohjain, joka kykeni suorittamaan koko grafiikkaliukuhinnan yhdessä paketissa hyödyntäen Unified shader arkkitehtuuria. [27] Näin syntyivät ytimet, joista kerrotaan luvussa 5.2 lisää.

GeForce 8800 oli myös ensimmäinen näytönohjain, joka kykeni hyödyntämään CUDA-ohjelmointiympäristöä, sillä Unified shader arkkitehtuuri mahdollisti ytimien käyttämisen yleisiin laskutoimituksiin, eli muuhunkin kuin grafiikan renderöimiseen. Tämä oli GPGPU:n ("General Purpose Graphics Processing Unit") ja CUDA-ajan alku.

Muisti oli alkuaikoina paljon rajoittuneempaa ja vasta monia sukupolvia myöhemmin muistien ominaisuudet tiedonsiirron ja kapasiteettien ulkopuolelta paranivat. Yleisesti näytönohjaimien muisti kuitenkin toimii ikään kuin rinnakkain, eli prosessori pystyy ottamaan muistista ”viereiset palikat” niiden ollessa sijoitettuna muistiin peräkkäin. Näytönohjaimien muisti on siis hajautettuina lohkoihin, joista voidaan hakea keskenään samaan aikaan dataa. Muistin suunnittelussa onkin painotettu kaistanleveyttä viiveen sijaan. Moderneilla näytönohjaimilla muistin kaistanleveys voi vaihdella mallin mukaan ~200–1000 GB/s välillä. [27] GeForce 256:lla muistin kaistanleveys oli suhteessa vain vaivaiset 2,66 GB/s [14]. Muistin kaistanleveys saadaan laskettua kertomalla muistin nopeus väylän leveydellä.

Taulukossa 1 on vertailu Nvidian neljästä viimeisimmästä peruskäyttöön suunnatusta näytönohjaimesta. Arkkitehtuurien vertailu on suoritettu huippumallien perusteella, sillä se antaa parhaan kuvan arkkitehtuurin kapasiteetista sekä ominaisuuksista ja vertailu olisi muuten hieman haastavaa muuttujien määrän kasvaessa. Huomataan että RTX 2080 Ti ja GTX 1660 Ti käyttävät samaa arkkitehtuuria mutta se on ominaisuuksiltaan kuitenkin paljon rajatumpi sen ollessa halpa ja matalan tason käytölle tarkoitettu näytönohjain. Lyhyesti sanottuna ominaisuudet ovat parantuneet jokaisella sukupolvella.

**Taulukko 1.** Vertailu Nvidian 4:stä viimeisimmistä myynnissä olleesta lippulaivamalleista [29].

Näytönohjain	RTX 4090	RTX 3090 Ti	RTX 2080 Ti	GTX 1660 Ti
Arkkitehtuuri	Ada Lovelace	Ampere	Turing	Turing
Ray Tracing-ytimet	Gen 3	Gen 2	Gen 1	
Tensor-ytimet	Gen 4	Gen 3	Gen 2	
CUDA-ytimien määrä	16 384	10 752	4352	1536
Muistin konfiguraatio	24GB GDDR6X	24GB GDDR6X	11GB GDDR6	6GB GDDR6
muistin väylän leveys	384-bit	384-bit	352-bit	192-bit

Verrattuna taulukossa 2 esiteltyyn GeForce 8800 GTX näytönohjaimeen, voidaan kehityksen todeta olleen suurta 18 vuoden aikana. Vaikkakin huomattavaa on, että vaikka näytönohjaimet ovat kehittyneet pitkälti jokaiselta aspektilta muuten paremmaksi, jää silti esimerkiksi GTX 1660 Ti jalkoihin muistin väylän leveyden suhteen. Väylän leveys ei kuitenkaan kerro koko tarinaa. GeForce 8800 GTX näytönohjaimen arkkitehtuurilla kaistanleveys on 86,4 GB/s [30] kun taas GTX 1660 Ti arkkitehtuurilla se on jopa 288,1 GB/s [31]. Vaikka ero on iso, kalpenee se silti erolle, joka on tapahtunut noin seitsemässä vuodessa GeForce 256:n GeForce 8800 GTX:n välillä. Kaistanleveydessä on tapahtunut yli 30 kertainen kasvu.

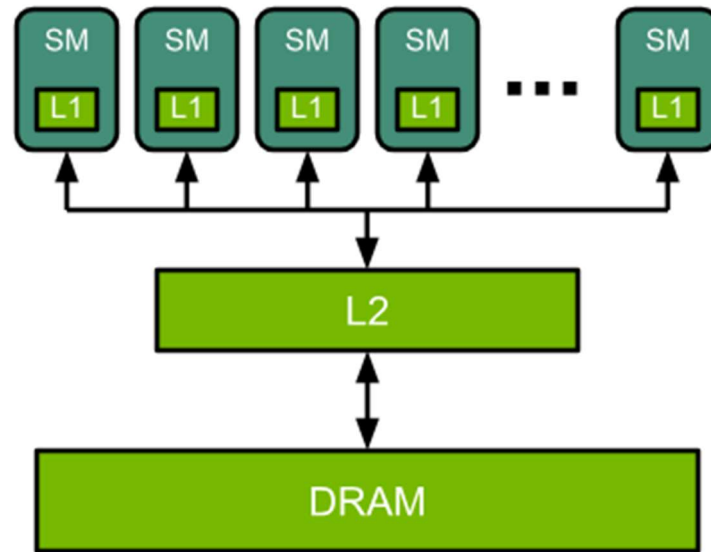
**Taulukko 2.** Nvidia GeForce 8800 GTX arkkitehtuurin ominaisuudet [30].

Arkkitehtuuri	Tesla
Ray Tracing-ytimet	
Tensor-ytimet	
CUDA-ytimien määrä	128
Muistin konfiguraatio	768MB GDDR3
Muistin väylän leveys	384-bit

## 5.2 Ydinten ja laskentatehon lisääntyminen

Keskusprosessoreja voidaan kuvata englanninkielisellä termillä *multicore processors* ja näytönohjaimia puolestaan *manycore processors*. [27] Tämä johtuu siitä, että moderneilla keskusprosessoreilla ytimiä on nykyisin usein kuudesta 24:ään, kun näytönohjaimilla niitä on tuhansista yli kymmeneen tuhanteen. Vasta vuodesta 2006 lähtien on voitu puhua virallisesti ydinten lukumääristä Nvidian kohdalla, sillä GeForce 8800 oli ensimmäinen näytönohjin, joka korvasi erilliset shader piirit ytimillä. Tämän takia ydinlukumäärävertailu Nvidian kohdalla on mahdollista aikaisintaan 8800 malleista nykypäivän näytönohjaimiin. AMD:lla puolestaan rajoitus tapahtuu toukokuussa 2007 julkaistusta Radeon HD 2900 XT näytönohjaimesta nykypäivään [32].

Nykypäivänä ytimet koostuvat useista osista ja ytimeistä itsestään käytetään nimitystä Streaming Multiprocessor (SM) tai (SMX). Kepler arkkitehtuuri yksinään käytti SMX-terminiä [33], mutta selkeyden puolesta jatkossa käytetään SM-terminiä niiden ollessa tarpeeksi samanlaisia, kun se on mahdollista. Yksinkertaisesti jokaisessa ytimessä on paljon CUDA-ytimiä, joiden vastuulla ovat primitiiviset laskutoimitukset. Eri arkkitehtuureilla on eri määrä CUDA-ytimiä jokaista SM-ydintä kohden. CUDA-ytimien rinnalta löydetään rekisteri, ns. register file (RegF), joka sisältää datan, jota CUDA-ytimet käsittelevät. Data rekisteriin siirretään prosessorin ulkopuolisesta DRAM-muistista (Dynamic Random-Access Memory). Rekisterimuistin lisäksi SM sisältää L1- ja L2 cache-muistia, jotka ovat järjestään pienempiä. Cache muisti on muistia, jota näytönohjin hallitsee itsenäisesti. [34][35] Kuvassa 6 on esiteltynä yksinkertainen näytönohjinprosessorin rakenne.



**Kuva 5.** Yksinkertainen GPU-rakenne [35].

Kuluneiden vuosien aikana nyrkkisääntönä voidaan pitää, että ytimien määrä ja täten laskentateho on kasvanut. Esimerkiksi RTX 4090:lla on 16384 CUDA-ydintä, RTX 2080 Ti:lla niitä on 4352 kappaletta ja 8800 GTX:lla puolestaan 128 kappaletta [29][30]. Hyyt ovat isoja, mutta huomioitaessa näytönohjainten julkaisuvedot, tuntuvat hyyt vielä isommilta. RTX 4090 julkaistiin vuonna 2022, RTX 2080 Ti vuonna 2018 ja 8800 GTX vuonna 2006. Kahden vertailussa uusimman näytönohjaimen välillä laskennallisesti ytimiä olisi tullut vuodessa lisää 3008 kappaletta, kun taas kahden vanhimman välillä vuodessa lisäys olisi ollut 352 ydintä per vuosi. On huomioitava, että todellisuudessa ydinten määrän kasvu ei ole lineaarinen ajan suhteen.

Laskentatehon kasvua voidaan tutkia usealla eri tavalla, kuten esimerkiksi kuinka monta ruutua sekunnissa näytönohjain pystyy tuottamaan sekunnissa (FPS, Frames Per Second) tai kuinka monta laskutoimitusta sekunnissa näytönohjain pystyy suorittamaan. Käyttötarkoituksen mukaan eri metriikat kertovat erilaisia asioita eivätkä suoraan kerro hirveän hyvin yleisestä suorituskyvystä suuressa kokonaiskuvassa. Nyt kuitenkin tyydytään FLOPS-metriikkaan Shader ytimien suhteen sen antaessa kohtuullisen suuntaa antavan kuvan laskutehosta. RTX 4090 kykenee huikeaan 83 TFLOPS:iin, RTX 2080 Ti noin 27 TFLOPS:iin ja 8800 GTX noin 0,52 TFLOPS:iin [29][36]. Voidaan todeta, että laskutehon nousu on selkeästi seurausta ytimien määrän kasvusta.

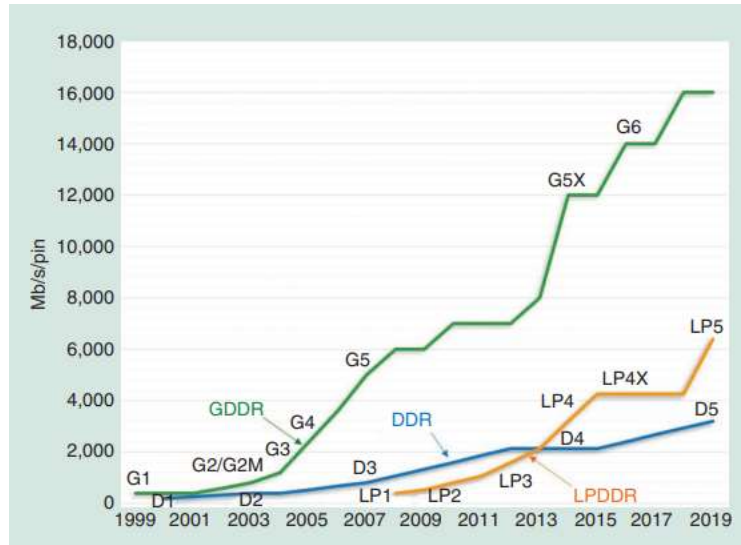
### 5.3 Muistin ja nopeuksien kasvaminen

Tässä aliluvussa RAM-muistilla viitataan näytönohjaimissa käytettävään GDDR-muististandardiin (Graphics Double Data Rate), joka pohjautuu DRAM-teknologiaan, ellei toisin mainita. Taulukossa 3 on esiteltyä GDDR-muistien julkaisuvuodet ensimmäisestä GDDR1-muistista GDDR7-muistiin, sillä kuva 6 on alun perin julkaistu ennen mainitsemisen arvoisia GDDR6X ja GDDR7 muisteja.

**Taulukko 3.** *GDDR-muististandardien julkaisuvuodet [37].*

Muisti	GDDR1	GDDR2	GDDR3	GDDR4	GDDR5	GDDR5X	GDDR6	GDDR6X	GDDR7
Vuosi	2000	2003	2005	2006	2008	2016	2018	2020	2023

DDR-muisti (Double Data Rate) itsessään on vuonna 2000 Joint Electron Device Engineerin Councilin (JEDEC) julkaisema standardi [38] joka mahdollistaa datan siirtämisen sekä nousevalla että laskevalla kellonreunalla. Esimerkiksi yleinen tietokoneissa käytettävä DDR4-RAM muisti, DDR4-3200 PC4-25600, pystyy suorittamaan 3200 miljoonaa tiedonsiirtoa sekunnissa eli MT/s (Millions of Transfers per second) huolimatta tiedonsiirtoväylän kellotaajuudesta, joka on vain puolet tästä, 1600 MHz. Tämä saavutetaan DDR-standardin ansiosta. Nimen toinen numerosarja tarkoittaa teoreettista muistin kaistanleveyttä, tässä tapauksessa se olisi 25,6 GB/s. [40] On huomioitava, että vaikka GDDR-muisti on myös Double Data Rate muistia, on sillä huomattavan paljon suuremmat siirtomäärät sekunnissa johtuen esimerkiksi leveämmistä tiedonsiirtoväylistä. Esimerkiksi yksi nopeimmista ja uusimmista DDR5 muisteista kykenee 6400 MT/s nopeuksiin, kun GDDR7 tulee kykenemään jopa 40 000 MT/s nopeuksiin. Jopa 16 vuotta vanha GDDR5 yltää 8000 MT/s nopeuksiin. Nämä korkeat nopeudet ovat kriittisiä näytönohjaimien suorituskyvyn kannalta ja viimeisten muutaman GDDR-muistisukupolven aikana on nähty todella nopeaa kehitystä, joka korreloi näytönohjainten laskentatehon nopean kasvun kanssa yhdessä näytönohjaimen ytimien ja arkkitehtuurin kanssa. Kokonaiskaistanleveys saadaan laskettua, tiedettäessä kuinka nopeasti tiedonsiirto tapahtuu (MT/s), siirrettyjen bittien määrä per siirto (DDR kohdalla kaksi bittiä) ja tiedonsiirtoväylän leveys bitteinä.



**Kuva 6.** GDDR1-6 muistien kaistanleveys yhtä pinniä kohden [38].

Kuten kuvasta 6 voidaan huomata, muistien kaistanleveydet pinniä kohden ovat kasvaneet vuosien saatossa ja täten ovat koko muistin kaistanleveydetkin. Vaikka kuvassa ei ole esiteltyä viimeisimpiä muistiteknologioita, on kehitys jatkunut näiden kohdalla yhtä lailla. Muistin nopeuden kasvun lisäksi toinen merkittävä tekijä muistin kehityksessä on kapasiteetin kasvu. Suurempi määrä muistia näytönohjaimen käytettävissä tarkoittaa suurempaa määrää dataa, joka voidaan varastoida kerralla muistiin. Tämä vähentää tarvetta ladata tasaisin väliajoin tietokoneen massamuistista dataa välimuistiin, joka voi varsinkin hitaammilla järjestelmillä aiheuttaa esimerkiksi videopelien pysähtymistä siksi aikaa, kunnes tarvittava uusi data saadaan ladattua käsiteltäväksi.

Kuvasta 6 voidaan myös huomata sinisellä viivalla kuvatun tavallisen RAM-muistina käytetyn DDR-muistin kaistanleveyksien pinniä kohden olevan huomattavan paljon pienempiä kuin GDDR-muistin. Tämä johtuu siitä, että DDR-muistissa priorisoidaan mahdollisimman pientä viivettä keskusprosessoria varten. Näytönohjaimet tarvitsevat puolestaan mahdollisimman paljon dataa kerralla pienemmän viiveen sijaan.

Kaikkien eri GDDR muistisukupolvien vertailu keskenään olisi epäkäytännöllistä tämän työn kokoluokassa, joten tyydytään vertailemaan nyt viimeaikojen yhtä yleisimmistä muistityypeistä GDDR6:sta uuteen GDDR7 muistiin. GDDR7 muistia oletetaan löytyvän seuraavan sukupolven RTX 5090 näytönohjaimesta vuonna 2025. Mainitsemisen arvoista on, että verrattaessa GDDR7 muistia ensimmäiseen GDDR muistisukupolveen, on tapahtunut 160-kertainen kasvu kaistanleveydessä pinniä kohden 25 vuodessa. Taulukossa 4 on esiteltyä näiden kahden muistin eroja.

Micronin mukaan, joka on yksi kolmesta muistivalmistajasta Samsungin ja Hynixin lisäksi ja jonka sivuilta datalehdeltä tieto on haettu, GDDR7 tulee saavuttamaan keskimäärin jopa yli 30 % kasvun FPS lukemissa verrattuna GDDR6/6X muisteihin. Lisäksi Micronin mukaan suuremman järjestelmänkaistanleveyden ansiosta tekoälyjen vastausaika tulee pienenemään jopa 20 % verran tekstistä kuvaksi prosesseissa. Lisäksi tehonkulutus tulee laskemaan hieman, noin 2 pikojoulea yhtä siirrettyä bittiä kohden.

**Taulukko 4.** GDDR6 ja GDDR7 vertailu [39].

Ominaisuus	GDDR6	GDDR7
DRAM sirun muistitiheys	16 Gb	16 Gb
Kaistanleveys pinniä kohden	18 Gbps	32 Gbps
Kaistanleveys komponenttia kohden	72 GBps	128 GBps
Kaistanleveys järjestelmää kohden	960 GBps	1,5 TBps
ECC	ei	kyllä
Teho per bitti	6,5 pJ/bit	4,5 pJ/bit

## 5.4 Lämmönhallinta ja virrankulutus

Laskentatehon kasvaessa, on normaalia, että tehonkulutus kasvaa ja täten myös syntyvä lämpöenergian määrä sekä virrankulutus. Liiallinen lämpöenergia aiheuttaa suorituskyvyn laskua, heikentää komponenttien luotettavuutta ja mahdollisesti lyhentää käyttöikä.

Näytönohjainten alkuaikoina riittävä jäähdytysteho saatiin aikaiseksi usein yhdellä mikroprosessorin päälle asetetulla tuulettimella ja pienellä jäähdytyslevyllä, kuten kuvassa 3 GeForce 256 kohdalla. Nykypäivinä tehokkaimmat näytönohjaimet vaativat jopa kolme tuuletinta ja massiivisen jäähdytyslevyn, joka on usein pidempi kuin itse näytönohjainkortti. Tämä voidaan nähdä kuvasta 7, jossa esitellyn RTX 4090:n jäähdytyslevy jatkuu pitkälle piirilevyn yli. Samalla on nyt hyvä huomioida, kuinka pieni näytönohjaimen piirilevy on suhteessa koko pakkauksen. Valtaosa näytönohjainkorttien koosta johtuu puhtaasti jäähdytysratkaisuista. Pienempitehoisilla näytönohjaimilla koko on myös pienempi niiden tuottaessa vähemmän lämpöä, jolloin riittää huomattavasti pienempi jäähdytyslevy. Kokoon vaikuttaa lisäksi tarvittavien komponenttien määrät ja koot, kuten GDDR-muistin konfiguraatio.

Nykypäivänä on usein myös tarpeellista jäähdyttää mm. muistikomponentteja suurien nopeuksien tuottaman lämmön takia. Näille kuitenkin usein riittää fyysinen kontakti jäähdytyslevyn kanssa.



**Kuva 7.** Osittain purettu Asuksen valmistama Nvidia RTX 4090 STRIX OC näyttönohjain [42].

Ilmajäähdytys on kuitenkin vain yksi monista jäähdytysratkaisuista. Muita ratkaisuja voidaan kategorisoida kahteen eri luokkaan. Fyysisiä ratkaisuja ovat esimerkiksi nestejäähdytys, termosähköiset menetelmät tai olomuotoa muuttavat materiaalit muiden menetelmien tukena. Toinen vaihtoehto on muuttaa näyttönohjaimen toimintaa sähköisellä tasolla. Yksi käytetyimmistä menetelmistä on DVFS (Dynamic Voltage and Frequency Scaling), joka tarkoittaa, että prosessori voi säätää tarvitsemaansa jännitettä ja toiminta-aajuutta kuormituksen mukaan. [41] Tällöin selatessa esimerkiksi uutisia näyttönohjain voi käydä lähes minimitehoilla tuottaen hyvin vähän lämpöä, kun taas pelatessa se voi ottaa maksimitehot ja saavuttaa täten parhaan suorituskyvyn tuottaen huomattavasti enemmän lämpöä. Usein näyttönohjaimien tapauksessa esimerkiksi pelien kehittäjillä on myös vastuu ja kyky hallita lämmöntuotantoa. Hyvin optimoitu ohjelmisto/peli vaatii vähemmän resursseja tuottaen vähemmän lämpöä.

Kuten luvun alussa mainittiin, virtojen määrät ovat kasvaneet sillä lämpömäärätkin ovat. Esimerkiksi GeForce 8800 GTX näyttönohjaimen maksimitehokulutus oli 155 W, kun taas kuvan 7 näyttönohjaimella tehonkulutus on parhaimmillaan noin 450 W. Tehonkasvu johdetaan kasvaneen virrantarpeesta, kun ytimiä on tullut lisää ja niiden kehittyessä niille on voitu syöttää enemmän ja enemmän virtaa, jotta saataisiin nostettua suorituskykyä jokaisen uuden sukupolven kohdalla. Tämän takia useat hyvin pienitehoiset tai vanhat näyttönohjaimet kykenivät saamaan virtansa suoraan emolevyllä, kun uudet ja korkeatehoiset näyttönohjaimet tarvitsevat erilliset virtajohdot suoraan tehonlähteeltä.

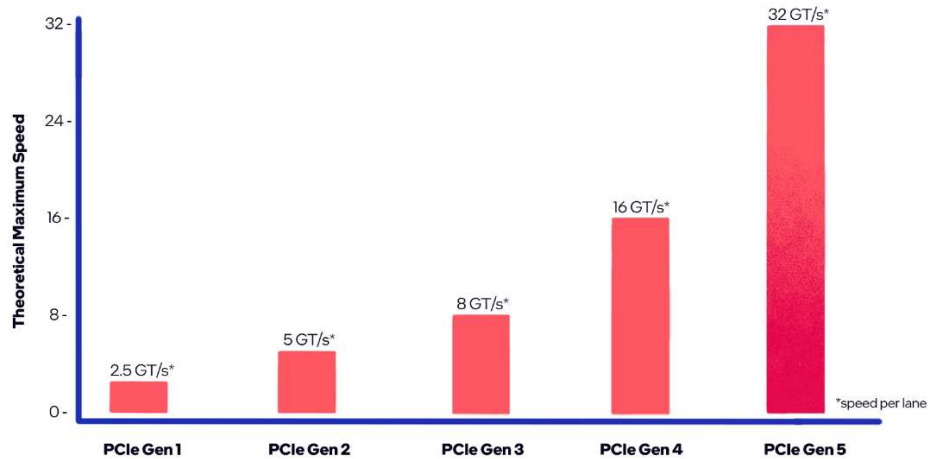
Kokonaisuudessaan lämmönhallinta on pysynyt pääosin samanlaisena vuosien ajan, ilmajäähdytys on ollut hyvä kompromissi luotettavuuden, toimivuuden ja hinnan kannalta. DVFS tai sen kaltaisia menetelmiä on ollut käytössä myös GeForce 8800 ajoilta alkaen.

## 5.5 Liitännät

Läpi historian näytönohjaimet ovat nähneet lukuisia eri liittimiä, joilla on ollut eri tarkoituksia; Ulostuloliittimet, komponenttien väliset I/O-väylät (Input/Output), rinnankytkentäliittimet ja virtaliittimet. Ulostuloliittimillä tarkoitetaan esimerkiksi nykypäivän näytönohjaimista löytyviä HDMI- ja DisplayPort-liittimiä. Vanhempia vaihtoehtoja ovat olleet esimerkiksi DVI-liittimet. Näiden kehityksen on määrännyt lähinnä näyttötekniikan kehittyminen, joten ne jätetään tarkastelun ulkopuolelle. Virtaliittimet ovat ainoastaan kortteja varten, jotka eivät kykene saamaan riittävästi virtaa I/O-väylien kautta, joten ne jätetään myös tarkastelun ulkopuolelle. Tarkasteltaviksi jäävät siis I/O-väyläliittimet ja rinnankytkentäliittimet.

I/O-väylät ovat liitännäportteja ja -liittimiä, joiden avulla komponentit kuten näytönohjaimet voivat viestiä keskusprosessorin kanssa. Nykyisin standardisoitunut PCIe väylä on käytetyin I/O-implementaatio. Intel julkaisi vuonna 2001 PCIe väylän ensimmäisen version, joka korvasi vuonna -92 julkaistun edeltäjänsä PCI-väylän sekä AGP-väylän [43], jota esimerkiksi GeForce 256 käytti [5]. AGP-portti otti vastaan ainoastaan AGP-liittimen, mutta AGP-liitin saatettiin kytkeä PCI-porttiin.

PCIe väylistä on olemassa useampia eri sukupolvia ja kokoja. Liittimet koot on luokiteltu seuraavasti; x1, x2, x4, x8, x16. Mitä suurempi koko on, sitä enemmän se tarjoaa dataväyliä ja täten suuremman kaistanleveyden. Näytönohjaimet kytketään yleensä ylimpään x16 PCIe väylään sen ollessa lähimpänä suoritinta tarjoten täten parhaan nopeuden ja kaistanleveyden. Puolestaan jokainen uusi sukupolvi on ollut karkeasti tuplasti niin nopea kuin edeltäjänsä, kuten kuvasta 8 voidaan huomata. Uudet sukupolvet tukevat myös vanhempia sukupolvia. Esimerkiksi PCIe 3.0 laite voidaan liittää PCIe 5.0 porttiin. [44]



**Kuva 8.** PCIe väylien teoreettiset maksiminopeudet [44].

Vaikka nopeudet ovat kasvaneet jatkuvasti, Intelin mukaan ero suorituskyvyssä FPS metriikalla näytönohjaimien kohdalla on reaali maailmassa hyvin pieni jo kolmannen ja neljännen PCIe-sukupolven välillä. Joidenkin testien mukaan ilmeisesti viimeisimmät näytönohjaimet eivät saa käytettyä koko PCIe 3.0 x16 väylää edes 4K-resoluutiolla. Pelit voivat latautua nopeammin, kun näytönohjaimen videomuistiin saadaan siirrettyä tietoa nopeammin, mutta FPS-lukuihin se ei juurikaan vaikuta, vaan lähinnä latausaikoihin. [44]

Rinnankytkentäliittimet ovat kytkentäliittimiä, joiden tarkoituksena on liittää näytönohjaimia toisiinsa näytönohjainmuistin kasvattamiseksi ja kuorman jakamiseksi näytönohjaimien kesken, jolloin laskentatehoa saadaan kasvatettua. Esimerkiksi tunnetuin ja ensimmäinen metodi kytkentätavasta on SLI tai koko nimeltään NVLink-SLI. Ensimmäinen SLI-teknologiaa käyttänyt näytönohjain oli jo mainittu GeForce 6800 [24]. SLI oli pitkään perinteisesti käytetty standardi graafisissa käyttötarkoituksissa, kunnes Turing-arkkitehtuurissa sen korvasi NVLink-V2:een pohjautuva liitäntäteknologia. NVLink-V2:sta ja tätä edeltänyttä NVLink-V1:sta käytettiin lähinnä ammattitason näytönohjaimilla, joita voidaan kytkeä jopa 8 tai 6 rinnan riippuen teknologiasta. SLI loi pohjan NVLink teknologialle. [45] SLI mahdollisti kuvan tuomisen näytölle yhteensä viidellä eri tavalla. Näistä viidestä esimerkiksi AFR (Alternate Frame Rendering) jakoi kuorman tasan näytönohjaimien kesken. Kahden näytönohjaimen tapauksessa ne renderöivät kuvan vuorotellen. SFR (Split Frame Rendering) puolestaan jakoi kahden näytönohjaimen tapauksessa renderöitävän kuvan kahteen osa-alueeseen, jolloin molemmat näytönohjaimet työstivät samaa kuvaa. Renderöintiä tasapainotettiin dynaamisesti SFR:n tapauksessa, mikäli järjestelmä huomasi toisen näytönohjaimen kuormittuvan toista enemmän. Kolmantena mainitsemisen arvoisena menetelmänä oli niin kutsuttu Boost Performance Hybrid SLI

rendering. Tässä konfiguraatiossa toinen näytönohjaimista oli todennäköisesti aina tehokkaampi ja sen rinnalla oli pienempitehoinen näytönohjain. Käytännössä tämä menetelmä toimi kuten AFR, mutta kuorma oli painotettu tehokkaamman näytönohjaimen puolelle. [46] SLI kuitenkin vaipui pelikäytössä unholaan ja sen tuki onkin uudempien perusnäytönohjaimien kohdalla lakkautettu. Tähän on useita syitä, mutta käytännössä syynä oli tuplaantunut tehonkulutus ja vain marginaalinen suorituskyvyn nousu suurimmassa osassa videopelejä. Välillä jopa yksittäinen näytönohjain saattoi suoriutua paremmin verrattuna kahteen keskenään SLI-kytkettyyn. Tämä johtui pitkälti kuitenkin pelien optimoinnista yksittäisille näytönohjaimille, sillä vain murto-osa kuluttajista omisti kahta samaan aikaan käytössä olevaa näytönohjainta.

## 6. NÄYTÖNOHJAINTEN KÄYTTÖ PELAAMISEN ULKOPUOLELLA

Pelaamisen lisäksi näytönohjaimet ovat erityisesti nykypäivänä monen teollisuudenalan sekä käyttötarkoituksen kulmakivinä. Ilman näytönohjaimien mahdollistamaa erittäin korkeaa laskentatehoa, olisi hyvin moni asia toisin. Tässä luvussa käsitellään nyt missä ja miten näytönohjaimia käytetään pelaamisen ja kuvan tuottamisen lisäksi.

### 6.1 GPU datatieteessä, tekoälyssä ja laskennallisessa tieteessä

Datatieteessä, tekoälylaskennassa ja laskennallisissa tieteissä käsitellään usein erittäin suuria määriä dataa, joiden käsittely on aikaa vievää ja resurssi-intensiivistä. Tämän takia näytönohjaimia on aloitettu hyödyntämään paljon, sillä ne ovat suhteessa laskentatehoon, skaalattavuuteen sekä saatavuuteen nähden edullinen ratkaisu verrattuna CPU-laskentaan. Näytönohjaimia on kyetty hyödyntämään vuoden 2006 jälkeen Nvidian tuotua markkinoille Unified Shader arkkitehtuurin ja CUDA-ytimet, mahdollistaen näytönohjainten käytön muuhunkin kuin graafiseen laskemiseen. [47]

Suurien datamäärien käsittelyssä nykypäivänä auttavat koneoppimisalgoritmit, joiden pyörittämisessä näytönohjainten rinnakkaisuus on ratkaisevassa asemassa [47]. Teollisuudessa ja tieteessä käytetäänkin tehokkaita ammattikäyttöön tarkoitettuja näytönohjaimia, kuten Nvidia RTX A6000, jotka on suunniteltu palvelinkäyttöön ja käsittelemään suuria määriä dataa nopeammin ja tarkemmin kuin tavalliset pelikäyttöön suunnitellut mallit. RTX A6000 pohjautuu Ampere-arkkitehtuuriin ja tukee esimerkiksi kolmannen sukupolven NVLink-liitäntäteknologiaa mahdollistaen rinnankytkennän ja täten jopa 96GB näytönohjainmuistin, varta vasten suunnitellut CUDA-ytimet, jotka ovat kaksi kertaa energiatehokkaampia sekä suorittavat FP32-laskuja (Single-precision floating-point format) jopa tuplasti nopeammin kuin Turing-arkkitehtuuri [48]. Myös AMD-tarjoaa teollisuuskäyttöön suunnattuja Instinct-nimisiä näytönohjaimia erilaisella arkkitehtuurilla ja lähestymistavalla.

Eryityisesti modernit näytönohjaimet loistavat tekoälyn kouluttamisessa. Nvidian vuonna 2017 Volta-arkkitehtuurin mukana julkaisemat Tensor-ytimet ovat tekoälylaskentaan ja syväoppimiseen suunniteltuja ytimiä. Tensor ytimet ovat optimoituja matriisi- ja konvo-

luutiolaskentaan, mahdollistaen todella suuren suorituskyvyn niihin perustuvissa las-  
kuissa, joka on tärkeää tekoälyjen kouluttamisessa. Tensor ytimet ovat kehittyneet no-  
peasti lyhyessä ajassa, kuten kuvasta 9 voidaan havainnoida.

Generation	Architecture	Key Features	Performance Improvement
1st Gen	Volta	FP16 and FP32 support	Baseline
2nd Gen	Turing	INT8 and INT4 support	2x FP16 performance vs. Volta
3rd Gen	Ampere	TF32 format, sparse matrix support	5x FP16 performance vs. Volta
4th Gen	Hopper	FP8 support, Transformer Engine	6x FP16 performance vs. Ampere

**Kuva 9.** Tensor-ytimien kehitys menneiden sukupolvien välillä [49].

## 6.2 GPU:n rooli lohkoketjujen louhinnassa

Tunnetuimpana esimerkkinä lohkoketjuista ja niiden louhinnasta voidaan pitää kryptovaluuttojen louhintaa. Vuonna 2016 vallitsi kryptovaluuttaboomi, sillä näytönohjaimet mahdollistivat nopean lohkoketjujen louhinnan rinnakkaislaskennan ansiosta, joka johti kryptovaluuttojen arvon nousuun. Tämän seurauksena 2017 koettiin pulaa näytönohjaimista, sillä näytönohjaimia ostettiin suuria määriä sekä kryptovaluuttafarmeja varten kuten kuvassa 10 ja useat yksityiset henkilöt ostivat suuria määriä näytönohjaimia ja myivät niitä kalliimmalla. [50] Ajan kuluessa kryptovaluuttojen louhinta kävi kuitenkin hyvin työlääksi erityisesti isojen paljon louhittujen kryptovaluuttojen kohdalla. Tämän seurauksena alalla alettiin siirtyä yksinomaan kryptovaluuttojen louhintaan erikoistuneilla laitteilla. Kryptovaluuttojen louhinnan ulkopuolellakin näytönohjaimet ja lohkoketjut edustavat ideaalista yhdistelmää.



*Kuva 10. Kryptovaluuttojen louhinnassa käytetty kryptovaluuttafarmi [51].*

Suuremmissa kokonaiskuvassa lohkoketjut perustuvat prosessiin nimeltä Proof of Work (PoW), jonka avulla saadaan validoituja transaktiot ja suojeltua verkkoa. Tämä prosessi vaatii suurin määrän monimutkaisten laskutoimitusten ratkaisua, jossa näyttöohjaimet ovat hyviä kyetessään suorittamaan paljon laskuja samaan aikaan. Tulevaisuudessa lohkoketjujen merkitys tulee todennäköisesti kasvamaan ja täten tulee todennäköisesti näyttöohjaintenkin. [51]

### **6.3 Näyttöohjainten merkitys ja käyttö uusilla aloilla**

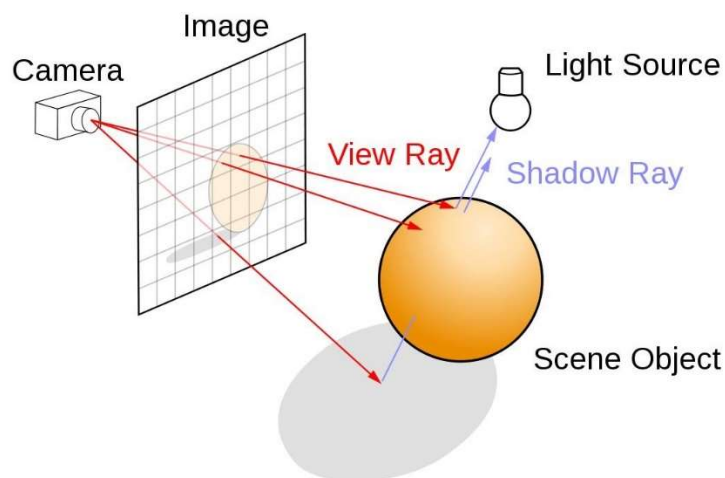
Näyttöohjaimia on alettu käyttämään yhä useammilla aloilla edellä mainittujen lisäksi. Kaikkia uusia käyttötarkoituksia yhdistävät kuitenkin pitkälti samat asiat kuin nykyisiäkin aloja. Näitä ovat erityisesti suuret datamäärät ja tarve niiden nopealle käsittelylle sekä tekoälyn integrointi. Tällaisia aloja ja käyttötarkoituksia ovat esimerkiksi itseohjautuvat autot ja niiden kehittäminen. Näyttöohjaimet tarjoavat suhteellisen edullisen keinon käsitellä dataa ja käyttää tekoälyä esimerkiksi uusien yritysten kohdalla, joilla ei välttämättä ole resursseja tai varallisuutta kehittää omaa järjestelmäänsä. Yksi merkittävä osa-alue missä näyttöohjainteknologiaa todennäköisesti hyödynnetään paljon, on sotilasteknologia. Esimerkiksi ohjusvalmistaja MBDA on julkaissut tavoitteensa saada Orchestrike niminen tekoäly- ja datalinkkijärjestelmä ohjuksiinsa, jotta ne voivat autonomisesti keskustella keskenään ja muuttaa iskustrategiaa vielä jopa 20 sekuntia ennen iskuhetkeä [52]. Näyttöohjaimet eivät ole kuitenkaan kaikkivoipia, jonka takia nykyisellä teknologialla on rajoituksensa sen suhteen mihin kaikkeen niitä voidaan käyttää. Näyttöohjain ei esimerkiksi pysty suorittamaan kaikkia asioita mitä keskusprosessori pystyy.

## 7. NYKYISET JA TULEVAISUUDEN TRENDIT

Trendit ovat olleet olennaisia kehityksen näkökulmasta jo pitkään, mutta viimeisen kymmenen vuoden aikana trendien luonne on alkanut muuttumaan merkittävästi. Puolestaan katsottaessa pitkälle tulevaisuuteen on vaikea ennustaa millaisia näyttöohjaimet ovat tulevaisuudessa ja miten niiden käyttömahdollisuudet ovat muuttuneet. Lähitulevaisuutta voidaan kuitenkin ennustaa hieman paremmin, kun tutkitaan mitä muuta uutta teknologiaa on kehitteillä jo tänään.

### 7.1 Ray Tracing ja muut teknologiat

Jo useampaan otteeseen mainittu säteenseuranta eli Ray Tracing, on yksi suurimmista nykypäivän trendeistä erityisesti peli- ja elokuva-alalla. Säteenseuranta on renderöintitekniikka, jonka avulla voidaan simuloida realistisesti valon käyttäytymistä, kuten varjoja sekä heijastuksia. Yksinkertaisuudessaan säteenseuranta tapahtuu seuraamalla valonsäteitä ns. virtuaalisesta kamerasta takaisin valonlähteeseen, kuten kuvassa 11 on esitetty. [53]



**Kuva 11.** Säteenseurannan periaate [53].

Luonnollisesti kuten jo paljon puhuttu, tekoäly on ehdottomasti yksi merkittävimmistä trendeissä näyttöohjainten näkökulmasta. Julkisesti saatavilla olevat tekoälymallit tarvitsevat käsittämättömän suuren määrän näyttöohjainsiruja, jotta niistä on saatu sellaisia kuin ne ovat nykyään. Tulevissa tekoälymalleissa lukumäärä tulee olemaan todennäköisesti vielä paljon suurempi. Jopa yksityiset henkilöt kykenevät nykypäivänä kouluttamaan alkeellisia omia tekoälyjään riittävän tehokkaalla tietokoneella sekä käyttämään

niitä paikallisesti ilman verkkoyhteyttä. Tekoälyä on yhdistetty jo paljon esimerkiksi pelaamisessa käytetyssä Nvidian DLSS- ja AMD:n FSR ruutugeneroinnissa.

Moderni virtuaalitodellisuus on ollut jo olemassa jonkin aikaa, mutta uusien tehokkaiden näytönohjainten myötä tälläkin saralla on otettu kuluneina vuosina huikeita loikkia. Esimerkiksi Applen Vision Pro VR-lasit mahdollistavat virtuaalitodellisuuden ja lisätyn todellisuuden (AR) yhdistämisen lähes saumattomasti. VR-laseissa voidaan käyttää myös DLSS- sekä FSR-ruutugenerointia suorituskyvyn parantamisessa. Vaikka nykyiset näytönohjaimet kykenevät pyörittämään VR-laseja ilman suurempia ongelmia, DLSS sekä FSR voivat tuoda entistä paremman kokemuksen tasaisemman ruudunpäivitysnopeuden ja keinotekoisesti korotetun resoluution avulla erityisesti hieman heikommille järjestelmille.

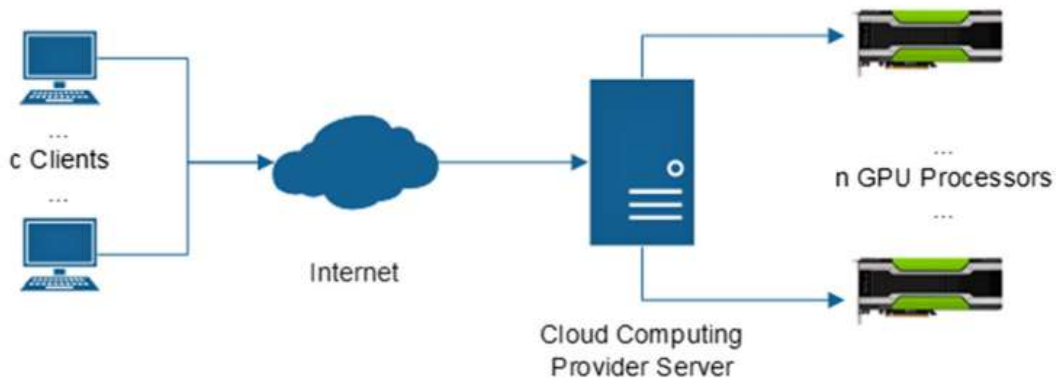
## 7.2 Pilvipohjainen grafiikan prosessointi

Pilvipohjaisessa grafiikan prosessoinnissa lähtökohtainen idea on, että raskas graafinen laskenta pelaamisessa tai teollisuudessa ulkoistetaan paikallisilta laitteilta isoille näytönohjainpalvelimille. Lisäksi pilvipohjainen laskenta mahdollistaa käytön virtuaalisesti mistä päin tahansa maailmaa. Kuvassa 12 on esiteltynä yksinkertainen pilvilaskennan periaate. Käytännössä data lähetetään palvelimelle, jossa näytönohjaimet suorittavat ohjeistetusti laskennan ja tämän jälkeen palvelin lähettää laskennan tulokset takaisin käyttäjälle.

Pelikäytössä graafisen prosessoinnin lisäksi kaikki muukin pelaamiseen tarvittava on sisällytetty palvelimille, jolloin käyttäjälle riittää riittävän nopea verkkoyhteys voidakseen nauttia pelistä kuin pelistä. Tämä mahdollistaa pelaamisen esimerkiksi hyvin edullisella kannettavalla tietokoneella, joka ei sisällä tarpeeksi tehokasta keskusprosessoria ja näytönohjainta pelien pyörittämiseen. Tällä hetkellä suurimpia verkkopelaamisen palvelua tarjoavat yritykset ovat Nvidia (Nvidia NOW), Microsoft (Xbox Cloud Gaming) sekä Sony (Playstation Cloud Streaming).

Pelaamisen ulkopuolella pilvipohjaista prosessointia hyödynnetään paljon mm. tekoälyjen kouluttamisessa, koneoppimisessa, tekoälypalvelun pyörittämisessä sekä datan käsittelyssä. Tällaisten palvelujen siirtäminen pilveen on myös taloudellisesti järkevä vaihtoehto yrityksille, sillä aloituskustannukset ovat murto-osa palvelun ostamisesta verrattuna paikallisen järjestelmän rakentamiseen. Lisäksi palveluntarjoaja saattaa mahdollistaa palvelimelta varatun laskentatehon muuttamisen tarpeen mukaan. Suuret palvelimet myös yleensä mahdollistavat todella paljon suuremman laskentatehon, joka nopeuttaa tuotantoa ja täten voi tuoda kilpailuetua palvelua käyttävälle yritykselle. Kuvassa 12 on

esiteltynä yksinkertainen pilvilaskennan periaate. Käytännössä data lähetetään palvelimelle, jossa näytönohjaimet suorittavat ohjeistetusti laskennan ja tämän jälkeen palvelin lähettää laskennan tulokset takaisin käyttäjälle.



*Kuva 12. Pilvilaskennan yksinkertainen periaate [54].*

### 7.3 Tulevaisuuden näköalat: kvanttilaskenta ja muut mahdollisuudet

Varsinkin viime aikoina on puhuttu paljon kvanttietokoneista ja niiden merkityksestä tietotekniikassa. Kun teknologia kypsyy joskus tulevaisuudessa, on oletettavaa, että näytönohjaimia tullaan integroimaan kvanttiteknoologiaan tai kvanttiteknoologiaa näytönohjainteknoologiaan. Näytönohjaimia voitaisiin käyttää mahdollisesti prosessoimaan dataa kvanttietokonetta varten valmiiksi tai jälkikäsittelemään dataa. Tällaisia käyttökohteita voisivat olla kryptografia sekä materiaalitieteet. [55]

Näytönohjaimet kuluttavat tunnetusti todella paljon energiaa suorittaessaan laskutoimituksia ja niiden sähköstä iso osa tuotetaan vielä fossiilisilla polttoaineilla. Puuttumatta sähköntuotannon menetelmiin, yksi vaihtoehto on kehittää arkkitehtuurista energiatehokkaampaa. [55] Esimerkiksi vuoden 2024 loppupuolella OpenAI:n toimitusjohtaja Sam Altman on tuonut Yhdysvaltain hallinnolle tarpeen ennennäkemättömän suurista datakeskuksista, jotka voisivat kulutta yhtä paljon energiaa kuin kokonaiset kaupungit. Nämä keskuksot tarvitsisivat noin 5 gigawatin tehomääriä per yksikkö, joka vastaa noin viittä ydinvoimalaa. Tällä energiamäärällä voisi tuottaa energiaa jopa melkein kolmelle miljoolle taloudelle. [56] Tällainen tehonkulutus on selvä merkki tarpeesta saada laskettua tehonkulutusta ilman, että laskentateho kärsisi merkittävästi. Tehonkulutus näkyy itse näytönohjaimien lisäksi datakeskuksissa esimerkiksi jäähdytysratkaisuihin.

Tulevaisuudessa tullaan todennäköisesti näkemään myös enemmän näyttöohjainteknologiaa IoT-laitteissa (Internet of Things). Tämä mahdollistaisi datan käsittelyn jo reaalitietokoneiden tasolla, eikä kaikkea tarvitsisi tehdä pilvessä. Tämä mahdollistaisi mm. tehokkaan tekoälyn integroimisen jo hyvin pienikokoisiin laitteisiin paikallisesti. Tämä vaatii kuitenkin näyttöohjainteknologian pienentymistä todella paljon, ellei sitä ennen onnistuta kehittämään pientä erillistä puhtaasti tekoälyyn erikoistunutta sirua. [55]

Käyttötarkoitusten lisäksi rakenteessa tulee todennäköisesti tapahtumaan muutoksia jo lähitulevaisuudessa, kuten mm. jo mainittu tehonkulutuksen minimoimisen tavoite. Yksi tavoite on siirtyä ns. heterogeeniseen arkkitehtuuriin, jossa näyttöohjaimen saataisiin integroitua jopa keskusprosessoreita ja FPGA-siruja (Field-Programmable Gate Arrays). Myös keskusprosessorin ja näyttöohjaimen muistia pyritään yhdistämään, jolloin voitaisiin välttyä monimutkaisilta tiedonsiirtomenetelmiltä ja saataisiin nopeutettua tiedonsiirtoa ja täten laskentaa. AMD:lla sekä Applella on käytössä jo osassa järjestelmiä yhdistettyjä muisteja, mutta ulkoisten irrallisten näyttöohjainten saralla näin pitkälle ei olla vielä päästy.

Tulevaisuuden ennustaminen kuitenkin todella vaikeaa, vaikka lähitulevaisuuden kehityksen suunnasta on jo vinkkejä. Pitkällä tähtäimellä on kuitenkin mahdollista ennustaa mihin suuntaan näyttöohjaimet ja niiden teknologia tulee kehittymään. Näyttöohjaimet voisivat joskus kaukana tulevaisuudessa korvata valtaosan tietokoneiden komponenteista tai muuttua puhelimen kokoisiksi palikoiksi tarjoten ennennäkemättömän laskentatehon verrattuna nykyisiin ratkaisuihin. Vain aika näyttää mitä tulevaisuus tuo tullessaan.

## 8. YHTEENVETO

Tässä kirjallisuustutkielmassa tutkittiin, miten irralliset näytönohjaimet ovat kehittyneet, mistä ne ovat saaneet alkunsa ja miltä niiden tulevaisuus näyttää.

Historiasta voidaan nostaa tärkeitä evoluution virstanpylväitä. Työssä käytettyjen materiaalien perusteella usea näytönohjainsukupolvi ja malli ovat nousseet esille useaan otteeseen. Näitä ovat esimerkiksi ensimmäinen virallinen näytönohjain, Nvidia GeForce 256, ensimmäiset kuvapuskurit, joista nykyiset näytönohjaimet saivat alkunsa, GeForce 8800-sarja ja CUDA-arkkitehtuuri, GeForce GTX 10xx-sarja, sekä GeForce RTX 20xx-sarja ja sen tuoma Ray Tracing. Näytönohjainten merkittävä tarjonnan kasvu 1990-luvulta alkaen on ehdottomasti merkittävin suoraan teknologian ulkopuolella oleva merkittävä tekijä, sillä se mahdollisti kehityksen jatkumisen. AMD:n kehittelemiä teknologioita ei tähän mahtunut mielestäni, sillä Nvidia on ollut pitkään pioneeri alalla ja tuonut valtaosan merkittävistä keksinnöistä kuluttajille ensin.

Ilman teknologista kehitystä alalla, nykypäivän yhteiskunta ei todennäköisesti olisi läheläkään niin kehittynyt kuin se on nyt. Lähes kaikki mitä teemme näyttöpäätelaitteilla vaativat toimiakseen näytönohjaimen ja kaikki mitä suunnittelemme vaativat näytönohjainteknologiaa. Pelaaminen on kehittynyt todella paljon ja tieteellinen laskenta vielä enemmän, kaikki kiitos näytönohjainten. On vaikea kuvitella nyky maailmaa ilman nykyisiä suuria laskentatehoja. Monet lääketieteelliset löydöt, tieteelliset läpimurrot, videopelit ja elokuvat olisivat jääneet saamatta.

Näytönohjainten rooli tulevaisuudessa tulee olemaan entistä merkittävämpi datamäärien kasvaessa ja tarpeiden monimutkaistuessa. Näytönohjaimien tulevaisuus itsessään on vielä mysteeri ja se tulee aina olemaan mysteeri, kukaan ei pysty ennustamaan millaisia näytönohjaimet tulevat olemaan esimerkiksi 20 vuoden päästä. Voimme vain arvailla katsomalla taaksepäin historiaan, miten näytönohjaimet ja niiden rooli tulevat muuttamaan ajan kuluessa.

## 9. LÄHTEET

- [1] Wikipedia Contributors. Tennis for Two. Wikipedia. Wikimedia Foundation; 2024 [viitattu 20.10.2024]. Saatavissa: [https://en.wikipedia.org/w/index.php?title=Tennis\\_for\\_Two&oldid=1251342280](https://en.wikipedia.org/w/index.php?title=Tennis_for_Two&oldid=1251342280)
- [2] Chris McClanahan. History and evolution of GPU architecture. Georgia Tech, College of Computing. [Viitattu 20.10.2024]. Saatavissa: <https://picture.iczhiku.com/resource/paper/WyiWoZOZoDZWgvxN.pdf>
- [3] Wikipedia Contributors. Graphics processing unit. Wikipedia. Wikimedia Foundation; 2024 [viitattu 20.10.2024]. Saatavissa: [https://en.wikipedia.org/wiki/Graphics\\_processing\\_unit](https://en.wikipedia.org/wiki/Graphics_processing_unit)
- [4] Dr. Jon Peddie. Famous Graphics Chips: Nvidia's GeForce 256. IEEE Computer Society. [Viitattu 20.10.2024, julkaistu 25.02.2021]. Saatavissa: <https://www.computer.org/publications/tech-news/chasing-pixels/nvidias-geforce-256>
- [5] Wikipedia Contributors. GeForce 256. Wikipedia. Wikimedia Foundation; 2024 [viitattu 20.10.2024]. Saatavissa: [https://en.wikipedia.org/wiki/GeForce\\_256](https://en.wikipedia.org/wiki/GeForce_256)
- [6] Wikipedia Contributors. Hercules Graphics Card. Wikipedia. Wikimedia Foundation; 2024. [Viitattu 23.10.2024]. [https://en.wikipedia.org/wiki/Hercules\\_Graphics\\_Card](https://en.wikipedia.org/wiki/Hercules_Graphics_Card)
- [7] Wikipedia Contributors. Color Graphics Adapter. Wikipedia. Wikimedia Foundation; 2024 [Viitattu 23.10.2024]. [https://en.wikipedia.org/wiki/Color\\_Graphics\\_Adapter](https://en.wikipedia.org/wiki/Color_Graphics_Adapter)
- [8] Centre for computing history. IBM Monochrome Display Adapter. [Viitattu 24.10.2024]. <https://www.computinghistory.org.uk/det/45038/IBM-Monochrome-Display-Adapter/>
- [9] Toy Story. IMDb. [Viitattu 18.11.2024]. <https://www.imdb.com/title/tt0114709/>
- [10] Henry E. Lowood. Doom. Britannica. [Viitattu 18.11.2024]. <https://www.britannica.com/topic/Doom>
- [11] Video Game History Timeline. The Strong National Museum of Play. [Viitattu 18.11.2024]. <https://www.museumofplay.org/video-game-history-timeline/>

- [12] Wikipedia Contributors. The Legend of Zelda: Ocarina of Time. Wikipedia. Wikimedia Foundation; 2024 [Viitattu 18.11.2024]. [https://en.wikipedia.org/wiki/The\\_Legend\\_of\\_Zelda:\\_Ocarina\\_of\\_Time](https://en.wikipedia.org/wiki/The_Legend_of_Zelda:_Ocarina_of_Time)
- [13] Web Design Museum. Walmart in 1999. [Viitattu 18.11.2024]. <https://www.web-designmuseum.org/gallery/walmart-in-1999>
- [14] Visiontek. 4x AGP GeForce 256 Graphics Accelerator. [Viitattu 18.11.2024]. [https://vgamuseum.info/images/doc/nvidia/gf256/geforce256\\_graphics.pdf](https://vgamuseum.info/images/doc/nvidia/gf256/geforce256_graphics.pdf)
- [15] A project of the Board of Regents of the University System of Georgia. A Brief History of the Internet, Sharing Resources. [Viitattu 23.11.2024]. [https://www.usg.edu/galileo/skills/unit07/inter-net07\\_02.phtml#:~:text=This%20allowed%20different%20kinds%20of,connected%20by%20a%20universal%20language](https://www.usg.edu/galileo/skills/unit07/inter-net07_02.phtml#:~:text=This%20allowed%20different%20kinds%20of,connected%20by%20a%20universal%20language)
- [16] The Editors of Encyclopedia Britannica. Erik Gregersen (viimeisin tarkistaja ja muokkaaja, 12.5.2021). Microcomputer. Britannica. [Viitattu 23.11.2024]. <https://www.britannica.com/technology/microcomputer>
- [17] Isaac V. Kerlov. The Art of Computer Animation and Effects, 4. painos, 2009. [Viitattu 23.11.2024]. Saatavissa: [https://books.google.fi/books?hl=en&lr=&id=7V87CwAAQ-BAJ&oi=fnd&pg=PR10&dq=1990%27s+3D+graphics&ots=8u4vHIDIA4&sig=8dd4bAmBFVEn56ZOZqGbyrlp-V8&redir\\_esc=y#v=onepage&q&f=true](https://books.google.fi/books?hl=en&lr=&id=7V87CwAAQ-BAJ&oi=fnd&pg=PR10&dq=1990%27s+3D+graphics&ots=8u4vHIDIA4&sig=8dd4bAmBFVEn56ZOZqGbyrlp-V8&redir_esc=y#v=onepage&q&f=true)
- [18] Antov Shilov. Tom's Hardware. Nvidia's grasp of desktop GPU market balloons to 88% — AMD has just 12%, Intel negligible, says JPR. [Julkaistu 6.6.2024] [Viitattu 28.11.2024] <https://www.tomshardware.com/pc-components/gpus/nvidias-grasp-of-desktop-gpu-market-balloons-to-88-amd-has-just-12-intel-negligible-says-jpr>
- [19] Wikipedia Contributors. List of defunct graphics chips and card companies. Wikipedia. Wikimedia Foundation; 2024 [Viitattu 28.11.2024]. [https://en.wikipedia.org/wiki/List\\_of\\_defunct\\_graphics\\_chips\\_and\\_card\\_companies](https://en.wikipedia.org/wiki/List_of_defunct_graphics_chips_and_card_companies)
- [20] John Wang, Jeffrey Hsu, Zhaoqiong Qin. A Comprehensive Analysis of Nvidia's Technological Innovations, Market Strategies, and Future Prospects. International Journal of Information Technologies and Systems Approach. DOI: [10.4018/IJITSA.344423](https://doi.org/10.4018/IJITSA.344423). [Viitattu 28.11.2024] Saatavissa: [https://www.researchgate.net/publication/380872958\\_A\\_Comprehensive\\_Analysis\\_of\\_Nvidia's\\_Technological\\_Innovations\\_Market\\_Strategies\\_and\\_Future\\_Prospects](https://www.researchgate.net/publication/380872958_A_Comprehensive_Analysis_of_Nvidia's_Technological_Innovations_Market_Strategies_and_Future_Prospects)

- [21] Ian. Nvidia and AMD: A Comparison of Innovations and Market Influence. [Julkaistu 27.06.2024] [Viitattu 29.11.2024] <https://press.farm/nvidia-and-amd-a-comparison-of-innovations/>
- [22] AMD Investor Relations, Press Release. AMD Unveils Next-Generation PC Gaming with AMD Radeon™ RX 6000 Series – Bringing Leadership 4K Resolution Performance to AAA Gaming. [Julkaistu 28.10.2020] [Viitattu 29.11.2024] <https://ir.amd.com/news-events/press-releases/detail/978/amd-unveils-next-generation-pc-gaming-with-amd-radeon-rx>
- [23] Nick Evanson. ATI Technologies: Gone But Not Forgotten. [Julkaistu 20.06.2023] [Viitattu 29.11.2024] <https://www.techspot.com/article/2689-ati-technologies-history/>
- [24] Nick Evanson. 25 Years Later: A Brief Analysis of GPU Processing Efficiency. TechSpot. [Julkaistu 14.04.2020] [Viitattu 30.11.2024] <https://www.techspot.com/article/2177-top-10-nvidia-graphics/>
- [25] Sami Haj-Assaad. Top 10 Most Significant Nvidia GPUs of All Time. TechSpot. [Julkaistu 23.12.2021] [Viitattu 30.11.2024] <https://www.techspot.com/article/2008-gpu-efficiency-historical-analysis/>
- [26] Sami Haj-Assaad. Top 10 Most Significant AMD GPUs of All Time. TechSpot. [Julkaistu 03.07.2021] [Viitattu 30.11.2024] <https://www.techspot.com/article/2172-top-10-amd-graphics/>
- [27] Mišić, Marko J. Durdević, Dorde M. Tomašević, Milo V. Evolution and trends in GPU computing. Scopus. [Julkaistu 2012] [Viitattu 09.12.2024] <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865088423&partnerID=40&md5=01f3beaeaa570301191beddb5c5e1920>
- [28] Nvidia. Vertex Shaders. [Viitattu 09.12.2024]. <https://www.nvidia.com/en-us/drivers/feature-vertexshader/#:~:text=A%20vertex%20shader%20is%20a,on%20the%20objects'%20vertex%20data>
- [29] Nvidia. Compare GeForce Graphics Cards. [Viitattu 10.12.2024]. <https://www.nvidia.com/en-us/geforce/graphics-cards/compare/>
- [30] Nvidia. Technical Brief, NVIDIA GeForce 8800 GPU Architecture. [Julkaistu 11.2006] [Viitattu 10.12.2024]. [https://www.nvidia.co.uk/content/PDF/GeForce 8800/GeForce 8800 GPU Architecture Technical Brief.pdf](https://www.nvidia.co.uk/content/PDF/GeForce%208800/GeForce%208800%20GPU%20Architecture%20Technical%20Brief.pdf)

- [31] Andrew Burnes. Nvidia. GeForce GTX 1660 Ti's Advanced Shaders Accelerate Performance In The Latest Games. [Julkaistu 25.02.2019] [Viitattu 10.12.2024]. <https://nvidia.com/en-us/geforce/news/geforce-gtx-1660-ti-advanced-shaders-streaming-multiprocessor/>
- [32] R. E. Poore, GPU-accelerated time-domain circuit simulation. 2009 IEEE Custom Integrated Circuits Conference. [Julkaistu 09.10.2009] [Viitattu 11.12.2024] <https://ieeexplore-ieee-org.libproxy.tuni.fi/stamp/stamp.jsp?tp=&arnumber=5280743>
- [33] Nvidia. Nvidia Kepler GK110 Next-Generation CUDA Compute Architecture. [Viitattu 11.12.2024]. <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/documents/NV-DS-Tesla-KCompute-Arch-May-2012-LR.pdf>
- [34] Xuan Shi, Volodymyr Kindratenko, Chaowei Yang. Modern Accelerator Technologies for Geographic Information Science. Springer. [Julkaistu 11.12.2024]. [https://link-springer-com.libproxy.tuni.fi/chapter/10.1007/978-1-4614-8745-6\\_1](https://link-springer-com.libproxy.tuni.fi/chapter/10.1007/978-1-4614-8745-6_1)
- [35] Nvidia. GPU Performance Background User's Guide. Nvidia Docs Hub. [Viitattu 11.12.2024]. <https://docs.nvidia.com/deeplearning/performance/dl-performance-gpu-background/index.html>
- [36] William Harmon. NVIDIA GeForce RTX 2080 Ti Graphics Card Review. STH. [Julkaistu 27.02.2019] [Viitattu 11.12.2024]. <https://www.servethehome.com/nvidia-geforce-rtx-2080-ti-graphics-card-review/3/>
- [37] Micron Technology. The evolution of GDDR: From GDDR1 to GDDR7. [Julkaistu 12.2024] [Viitattu 11.12.2024]. <https://www.micron.com/about/blog/memory/dram/the-evolution-of-gddr-from-gddr1-to-gddr7>
- [38] T. M. Hollis et al., "Recent Evolution in the DRAM Interface: Mile-Markers Along Memory Lane," in IEEE Solid-State Circuits Magazine, vol. 11, no. 2, pp. 14-30. [Julkaistu 24.06.2019] [Viitattu 11.12.2024]. <https://ieeexplore-ieee-org.libproxy.tuni.fi/stamp/stamp.jsp?tp=&arnumber=8741253>
- [39] Micron Technology. MICRON GDDR7 memory. [Revisio 06.2024] [Viitattu 11.12.2024]. <https://www.micron.com/content/dam/micron/global/public/products/product-flyer/gddr7-product-brief.pdf>
- [40] Wikipedia Contributors. Double data rate. Wikipedia. Wikimedia Foundation; 2024 [Viitattu 12.12.2024]. [https://en.wikipedia.org/wiki/Double\\_data\\_rate](https://en.wikipedia.org/wiki/Double_data_rate)

- [41] Hashemi Rafsanjani, Arpan Marwazi, Zoshim Olyver Wijaya T.Gari, Januar Tua Sinaga, Dahlan Sitompul. Thermal Management and Power Optimization in Modern CPU and GPU Architectures. Journal of Energy Engineering. [Julkaistu 06.12.2024] [Viitattu 12.12.2024]. [https://www.researchgate.net/publication/386476458\\_Thermal\\_Management\\_and\\_Power\\_Optimization\\_in\\_Modern\\_CPU\\_and\\_GPU\\_Architectures](https://www.researchgate.net/publication/386476458_Thermal_Management_and_Power_Optimization_in_Modern_CPU_and_GPU_Architectures)
- [42] W1zzard. ASUS GeForce RTX 4090 STRIX OC Review. TECHPOWEUP. [Viitattu 12.12.2024]. <https://www.techpowerup.com/review/asus-geforce-rtx-4090-strix-oc/3.html>
- [43] Xiao Qi Chen, Li yan Zhang, Chang Xian. FPGA-based Train Onboard PCIe Board Design and Implementation. Institute of Electrical and Electronics Engineers Inc. [Julkaistu 29.07.2023] [Viitattu 12.12.2024]. <https://ieeexplore-ieee-org.libproxy.tuni.fi/document/10262887>
- [44] Intel. What Are PCIe 4.0 and 5.0? [Viitattu 12.12.2024]. <https://www.intel.com/content/www/us/en/gaming/resources/what-is-pcie-4-and-why-does-it-matter.html#:~:text=Each%20generation%20of%20PCIe%20is,realized%20speeds%20may%20be%20slower>
- [45] A. Li et al., "Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect," in IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 1, pp. 94-110. [Julkaistu 01.01.2020] [Viitattu 12.12.2024]. <https://ieeexplore-ieee-org.libproxy.tuni.fi/stamp/stamp.jsp?tp=&arnumber=8763922>
- [46] Nvidia. SLI. [Viitattu 13.12.2024]. <https://docs.nvidia.com/gameworks/content/technologies/desktop/sli.htm>
- [47] Yiyu Cai, Simon See. GPU Computing and Applications. Springer. [Julkaistu 2015] [Viitattu 13.12.2024]. <https://link-springer-com.libproxy.tuni.fi/content/pdf/10.1007/978-981-287-134-3.pdf>
- [48] Nvidia. NVIDIA RTX A6000 datasheet. [Julkaistu 05.2022] [Viitattu 13.12.2024]. <https://resources.nvidia.com/en-us-briefcase-for-datasheets/proviz-print-nvidia-1?ncid=no-ncid>
- [49] Ravi Rao. Tensor Cores vs CUDA Cores: The Powerhouses of GPU Computing from Nvidia. Wevolver. [Julkaistu 25.06.2024] [Viitattu 13.12.2024]. <https://www.wevolver.com/article/tensor-cores-vs-cuda-cores>

- [50] Joe Liebkind. The GPU Industry Is Booming Thanks to Blockchain. [Päivitetty 24.10.2024] [Viitattu 13.12.2024]. <https://www.investopedia.com/tech/gpu-industry-booming-thanks-blockchain/>
- [51] Kevin Dwyer. Blockchain and GPU: Hardware Behind Decentralized Economies. Ankr. [Julkaistu 15.05.2024] [Viitattu 13.12.2024]. <https://www.ankr.com/blog/blockchain-and-gpu/>
- [52] John Hill. Orchestrike AI unveiled for the first time on SPEAR cruise missile. Airforce Technology. [Julkaistu 21.07.2024] [Viitattu 16.12.2024]. <https://www.airforce-technology.com/news/orchestrike-ai-unveiled-for-the-first-time-on-spear-cruise-missile/?cf-view>
- [53] Nvidia. Ray Tracing. [Viitattu 16.12.2024]. <https://developer.nvidia.com/discover/ray-tracing>
- [54] Víctor Sánchez-Ribes, Antonio Maciá-Lillo, Higinio Mora, Antonio Jimeno-Morenila. Efficient GPU Cloud architectures for outsourcing high-performance processing to the Cloud. Vol. 133, pp. 949–958. [Julkaistu 26.05.2023]. <https://link.springer.com/article/10.1007/s00170-023-11252-0>
- [55] Adrien Payong. Future Trends in GPU Technology. Digital Ocean. [Julkaistu 15.10.2024] [Viitattu 16.12.2024]. <https://www.digitialocean.com/community/conceptual-articles/future-trends-in-gpu-technology>
- [56] Bloomberg. OpenAI pitched White House on need of massive 5GW data center buildout. Business Standard. [Päivitetty 25.09.2024] [Viitattu 16.12.2024]. [https://www.business-standard.com/world-news/openai-pitched-white-house-on-need-of-massive-5gw-data-center-buildout-124092500102\\_1.html](https://www.business-standard.com/world-news/openai-pitched-white-house-on-need-of-massive-5gw-data-center-buildout-124092500102_1.html)