Industrifonden

2023-10-10

Investing in the Future of Computing

Iliam Barkino & Patrik Sobocki

Contents

About Industrifonden

Founded in 1979, Industrifonden has a proven legacy of backing emerging science and ground-breaking technologies so that they can transform into the industries of tomorrow. We manage more than SEK 5 billion in an evergreen fund, and invest in early-stage companies, from seed to series A, and sometimes earlier. Our first ticket is often 10-50 million SEK, with the capacity to invest considerably more during the lifecycle. Smaller tickets are also considered in an earlier stage for exceptional technologies. Our areas of expertise are Deep Tech, Life Science and Transformative Tech, and we currently have 50+ Nordic companies in our portfolio.

About the Authors

Iliam Barkino
Principal, Deep Tech

Iliam Barkino is a Principal within the Deep Tech practice at Industrifonden and has a passion for science and revolutionary technology. He holds a MSc in Engineering Physics, with focus on Computational Science. Iliam also studied business in parallel with his engineering studies and holds a BSc in Business and Economics. He has worked with large-scale AI transformations and conducted commercial due diligences at BCG, worked on numerous transactions at Industrifonden, and is on the board of three startups - including ZeroPoint Technologies in the Future of Computing space.

Patrik Sobocki
Senior Investment Director and Deep Tech Practice Lead

Patrik Sobocki leads the deep tech practice at Industrifonden with a specific interest in science-based technology in biology/healthcare and climate. Training in finance and PhD in data sciences turned serial entrepreneur. Over the past 10 years his focus has been on M&A in tech companies at growth stage in Europe and with Industrifonden for six years. Patrik is on the board of six startups - amongst other Enginzyme with a disruptive cell-free synbioplatform in biomanufacturing.

Executive summary

While there has been an enormous buzz about software, multiple trends are pushing the need for innovation to the hardware side of computing. These include the AI revolution, data explosion, and the memory bottleneck, to name a few. We explore some of the major hardware innovations within this field, both within existing computer architectures, and some that are based on other areas of science such as Bio-, Neuromorphic-, Optical-, and Quantum Computing. We also mention other advancements such as Ising machines and non-Neuromorphic in-memory computing.

We are interested in early-stage Nordic ventures with innovation across the computing space. However, this thesis focuses on hardware innovations and although it is not meant to provide comprehensive deep dives in any particular field of computing, we mention some sub-segments within the different computing fields that we are particularly interested in due to their potential impact, technology readiness, dependency of other innovations, and what we expect to come out of the innovation hubs in the Nordics.

In that spirit, we have a particularly strong interest in innovations that address the memory and interconnect within classical computing, optical interconnections due to their high impact and relative technology readiness, and ventures focusing on quantum middleware that can ride on previous hype waves around quantum hardware. Our view of the quantum field therefore remains the same as in 2021. We are also particularly interested in innovations around in-memory computing such as neuromorphic computing due to its potential impact, and hope that a recent surge in research and patents within the field will lead to an increase of ventures.

Introduction

The roots of modern computing trace back to the accomplishments of Ada Lovelace and Charles Babbage in the 19th century. Their collaborative efforts yielded the conceptual framework for what would later become the digital revolution. However, it wasn't until World War II that the potential of computing became imminent, with figures like Alan Turing and John Mauchly contributing significantly to the field's evolution in the purpose of decryption. Turing's pioneering work on codebreaking at Bletchley Park and the genius of the likes of Grace Hopper and her programming innovations laid the foundations for the electronic computing era.

In the 1940s, there was a need for more flexible programmable computing frameworks, which led to the introduction of the von Neumann architecture. It marked a pivotal moment and is characterized by the separation of the memory, which stores data and instructions, and the processing unit, which fetches data from the memory with the help of a bus and conducts operations according to the instructions. It has since then been the hallmark of classical computing and is still the most commonly used computer architecture, along with the Harvard architecture, which conceptually is the same but divides the memory into two: one for storing data, and one for storing instructions.

In later years, as computational demands have surged exponentially, the architectural framework that has served as the backbone of classical computing is reaching its limits. Once more we are coming to a pivotal moment with hardware innovations necessary to emerge and shape the next computational era. In this thesis we lay out some of the key trends that push the need for innovation within computing, explore some of the hardware developments addressing these trends, and share our view of investing in the future of computing.

Key trends shaping future of computing

Over the past years we have observed several macro trends that have increased the demand for higher computational power, such as the exponential progression of AI models and the explosion of data being generated. Prominent AI models today require 1,000,000x more operations to train than leading ones ten years ago. [2,3] At the same time, the development of computing power is getting restricted by memory access and the laws of physics. Businesses across industries also notice untapped value potential, available by solving complex optimization problems that are too computationally expensive for even the most powerful super computers available today to solve. Meanwhile, consumers expect their smart devices that come with more and increasingly complex sensors and functionalities to work seamlessly.

The following are some of the key trends that are shaping the future of computing. All of them, by themselves and collectively, require innovation on the hardware side of computing to meet the demand for faster, cost-effective, and environmentally-friendly solutions to store, transmit, and process data.

1. AI revolution

By now most of us have used large language models like ChatGPT at least once, and some have already incorporated it as a tool in their daily work and life. The AI revolution is marked by:

i) **The rapid adoption of AI technologies:** During the last decade AI has been at the forefront of reshaping industries, businesses, and everyday life. Businesses across all sectors have embraced AI as a strategic imperative, which has led to global AI adoption increasing from 20% in 2017 to 50% in 2022 while organizations have doubled the number of AI capabilities. [1]

ii) **Larger and more expensive models:** R&D within AI has seen a significant surge, driven by cutting-edge research. Over the span of five years, from 2016 to 2021, the volume of publications in this field nearly doubled. This activity has yielded not only a wealth of improved models that are larger, but also innovative use cases that have broadened the horizons of AI applications. [2] However, the potency of AI models has a flip side – their growing complexity. The race to create AI models with higher accuracy and greater capabilities has led to an exponential increase in model size. These behemoth models require formidable computing power for both training and inference, as posed by Exhibit 1. In addition, larger models require more time and energy to train and run – leading to significant environmental impact.

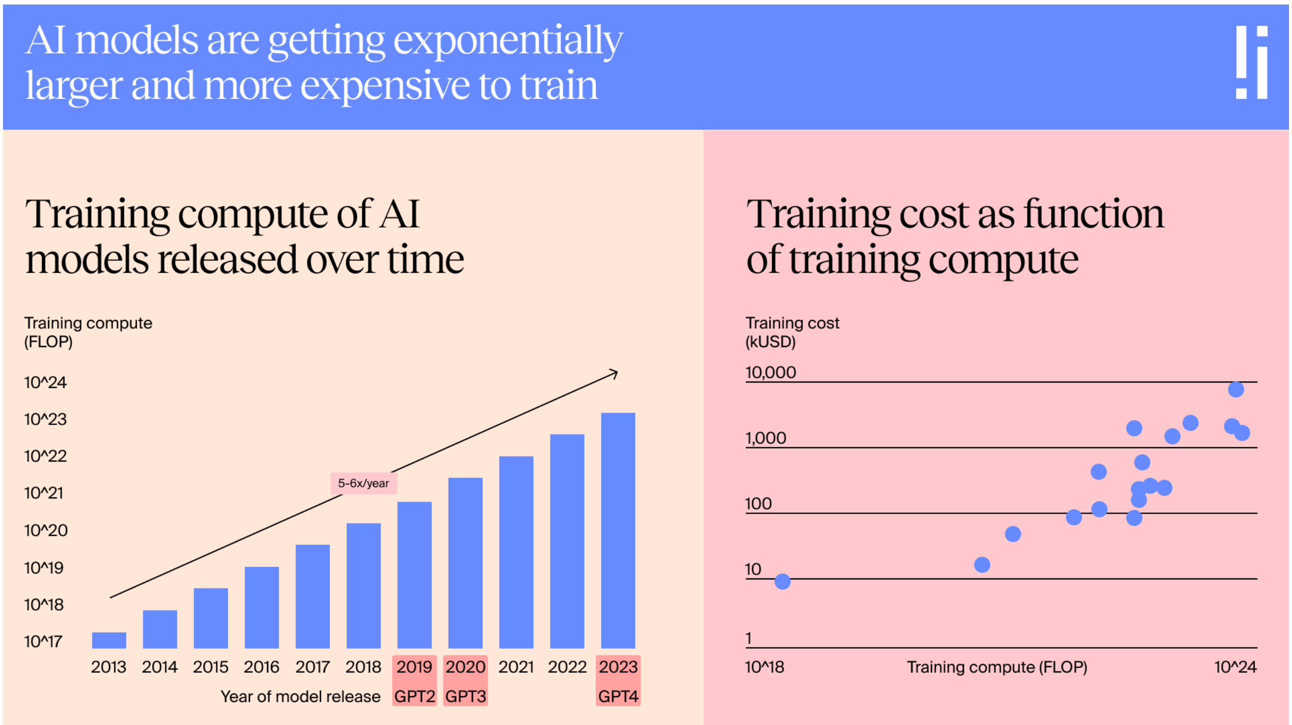

Exhibit 1. Sources: Stanford HAI [2], Sevilla et al. [3]

This has propelled computing requirement to new heights, demanding increasingly robust computational power to train and run models fast, at low cost, and with low environmental impact. The AI revolution thus acts as a litmus test for computing hardware, magnifying the need for innovation to cater to the intricate demands of AI workloads.

2. Data explosion

Over the past 10 years, businesses across sectors have embraced the digital transformation, with business leaders emphasizing the importance of “data-driven decisions”. A parallel trend that compounds the data landscape's evolution is the proliferation of data-generating devices. The advent of the Internet of Things (IoT) has seamlessly woven technology into the fabric of our daily lives, leading to connected devices increasing from from ~12Bn in 2014 to ~24Bn in 2022, and expected to reach ~46Bn by 2028. [4] From smart appliances and wearables to industrial sensors and autonomous vehicles, these interconnected devices generate a continuous stream of data - and regardless of if the data is sent to the cloud and data centres or processed on the edge, it must be managed somehow.

Alongside the increasing number of connected devices generating structured data from sensors, there is a projected growth rate of synthetic data generation that is at least three times faster than that of structured data. This synthetic data is primarily intended for training AI models, and this trend is expected to continue through 2030 according to Gartner research. By that time synthetic image- and video data will constitute more than 95% of data used for training AI models. [5]

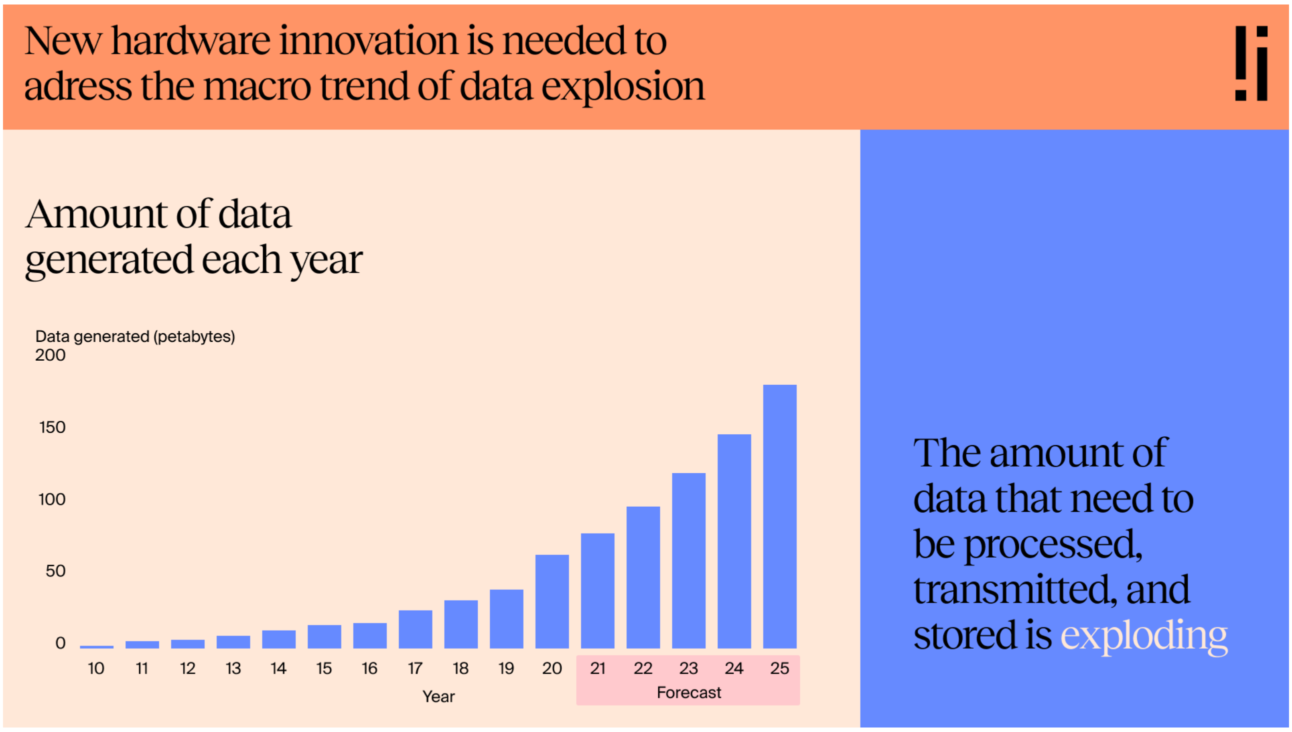

Exhibit 2. Source: Statista [6]

The combination of more data generating devices and more synthetic data generated has led to an explosive growth of data being generated, as presented by Exhibit 2, which is expected to continue grow exponentially in the near future. [6] These data need to be processed, analysed, and stored, which puts strains on memory and processing capabilities.

3. The memory bottleneck

Much of the focus of computing for the last years have been on the processing side, as newer chip generations have come with smaller transistors in larger quantities (see next trend) and more cores. More specialized processing units, primarily GPUs, have taken a larger chunk of the total computing tasks as they are more effective in executing specific types of computations highly parallelized, including graphic rendering and machine-learning training and inference.

However, as seen in Exhibit 3, the memory capabilities have not kept up with the speed at which processing power has increased. Memory bandwidth and DRAM frequency have both on average increased roughly 10% per year, while processor performance has increased almost 50% per year. [7,8,14]

Memory performance has not kept up with processor performance

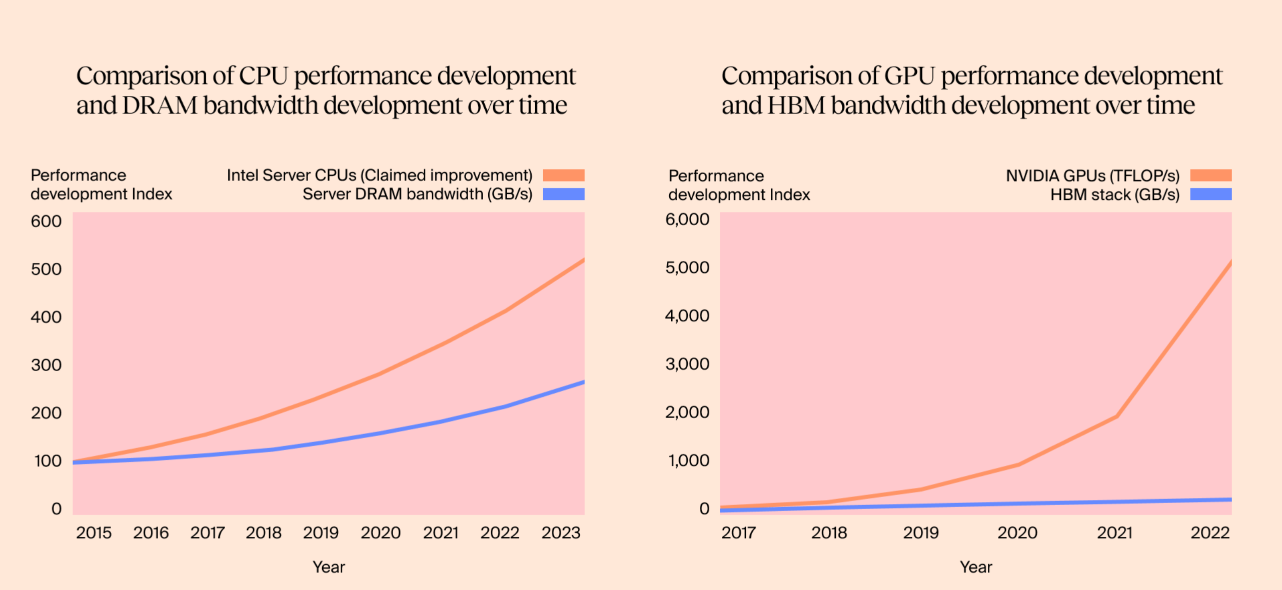

Exhibit 3. Sources: Intel [7], NVIDIA [8], Gartner [14]

The memory has thus become a bottleneck, even limiting processor performance. For example, it is not uncommon that high-performance processing units are idle 50% of the time, waiting for memory access. To quote a Gartner analyst "*We now see chips with teraflops scale processing performance, but only gigabyte scale data storage and interconnect.*" [14] This bottleneck requires that either the memory access is improved, or an architecture that reduces the need for reading and writing data to memory.

4. No Moore's law

Yes, this has been stated over and again in the field of computing. For the uninitiated, Gordon Moore, co-founder of Intel, predicted that the number of transistors that can be integrated onto a microchip would double every other year by shrinking the transistors, which would lead to corresponding improvements in overall computing performance. Transistors cannot continue to shrink indefinitely however and as they reach atomic scale, quantum effects begin to appear, making it difficult to maintain reliable operation and control the components, not to mention the increased cost and complexity of manufacturing, or the increased heat generation when transistors are more densely packed on a chip.

We will not go down the rabbit hole of arguing whether Moore's law is already dead by its original definition, slowing down, still holding true due to other innovation like multi-core processors, or whether it applies to specialized processing units like GPUs. Nonetheless we do argue that one of the most significant ways of increasing processor performance historically, namely shrinking transistors and adding more of them onto the chip, is reaching its natural peak. This trend urges for innovation to continue increasing processing performance.

5. Power consumption on the rise

With great computations comes great power consumption, and with great power consumption comes great responsibility. As chips become more complicated to manufacture, more data is produced, and AI-models become larger, the environmental impact of the ICT sector is rapidly increasing. Reports claim that the ICT sector's share of global energy demand can be up to 20% in 2030, 2-4x higher than in 2020, with data centres being one of the main contributors [9,10].

In addition to that, autonomous machines and devices that store and compute data locally need to be energy efficient in order to be practically useful. As energy demand increases with the complexity of the computing part of a device, energy-efficient alternatives for local compute are becoming more sought after.

6. Potential goldmines

Apart from the macro trends that emphasize the urgency for innovation within computing, there are multiple problems that, if solved, would tap into significant value potential across various industries. BCG estimated the value potential of solving these problems to over $450B. [15] Yet the problems are simply too hard for even the most sophisticated super computers to solve within reasonable time, as the number of variables in the optimization or number of possibilities a classical computer needs to explore quickly explodes. Below are a few such problems and why they would be valuable to solve.

i) **Advanced modelling:** Simulations of quantum mechanical systems, which would be useful to e.g. predict material properties or molecular interaction, which can accelerate design and development of new materials or drug molecules.

ii) **Combinatorial optimization:** Specific problems such as finding the shortest route for a delivery truck to all destinations.

iii) **Predictive analysis:** Modelling of complex systems, such as accurate long-term weather forecasting –which has high degree of dependency between variables and can help in predicting natural disasters, optimize crop management, etc.

Being able to solve such problems would be a goldmine of cost savings, time savings, and can be of utter strategic importance to homeland security.

Hardware developments addressing the computational gap

There are numerous areas within computing under development, to address the pain-points that arise with the macro trends described. In this section, we focus on some of the hardware developments and their potential impact. As to keep this report light touch, we do not deep dive into any of the areas here, but rather leave it for further discussion. We compare each of the initiatives with classical computing, i.e., von Neumann, and their use cases.

Accelerated computing (AC)

First out is what we call accelerated computing, which primarily involves innovations aimed at enhancing computing performance within existing classical architectures. While these innovations may not constitute a complete paradigm shift, they offer significant improvement in an established architecture. They include improvements on the processing, storing, and transmitting side of things.

On the processing side, NVIDIA's GPUs have so far dominated the field when it comes to AI applications, in addition to their original purpose in graphic cards. While they offer more flexibility and are more cost-effective for development purposes, we see a shift from general-purpose processors towards application specific integrated circuits (ASICs) within certain applications such as large-scale AI models. ASICS are however not programmable like e.g. FGPAs, and therefore need to yield significant performance improvements on large volumes of computations in order to add significant value. One notable startup in this field is Graphcore, a UK-based company developing an ASIC called the Intelligence Processing Unit (IPU), which promises to deliver a higher degree of parallelization. We anticipate that ASICs like the IPU will prove important in meeting the ever-increasing demand for computational power but will likely have to be somewhat reprogrammable for new AI applications since the AI field is constantly evolving.

We believe that hardware/software co-design, which involves the simultaneous design and optimization of both the hardware and software components of a system, are more likely to yield significant processing improvements. Cerebras, a US-based unicorn that has raised over €600M, have succeeded particularly in their hardware/software co-design approach with their deep learning-optimized chips CS-1 and CS-2. So, despite the hype around NVIDIA and ASICs, we focus more on startups with a hardware/software co-design approach on the processing side.

Additionally, we are particularly interested in startups addressing memory and interconnect issues. ZeroPoint Technologies is an example of such a company. They have developed a lossless, general purpose, hardware accelerated, compression algorithm that works across memory hierarchies – all the way from cache to storage. It deals with the issues of both memory capacity and memory bandwidth, yielding a 2-3x improvement in data compression and reducing processor idle time by 80% - thus addressing the bottlenecks created by current memory and interconnect performance.

There are of course other innovations within this field addressing these issues, such as high-bandwidth memory (HBM) or compute express link (CXL), just to name a few. HBM is simply put a 3D stacked DRAM which offers significant bandwidth improvements over DDR-based approaches but is currently also significantly more expensive than standard server DRAM. CXL is an emerging high-speed interconnect standard that holds great potential to improve memory capacity and bandwidth, but it is currently in moderately early stage of development.

Memory and interconnect bottlenecks remain the most pressing issue and solving these bottlenecks and has the highest potential impact within existing architectures, as they are currently limiting further processing improvements. Our investment focus within AC therefore remains on memory and interconnect innovation going forward but we stay receptive to potential innovations in processing-related technologies as well, especially if approached from a hardware/software co-design angle.

Bio computing (BC)

Although this is a very nascent field of computing that we believe is at least 10 years away, we want to briefly describe the potential of BC, which is the realm of processing, transmitting, and storing data with biological material and processes. There are a selected few projects on the processing side, e.g., Australian startup Cortical Labs that are developing a *Biological Intelligence Operating System* (biOS) by fusing lab-cultivated neurons from human stem cells onto computing devices. The company raised roughly USD 10M earlier in 2023.

Again, we focus more on the data storage side, namely DNA storage. In principle, DNA storage focuses on storing data in the form of DNA strands. Data is thus encoded into combination of A, G, C, T nucleotides which are synthesized as DNA strands, as opposed to electronical bits of 1 or 0. Retrieving data from DNA storage therefore necessitates sequencing the DNA and decoding the stored information. The three primary *theoretical* advantages of storing data in DNA include:

i) **High data density:** DNA can store magnitudes higher information per volume and weight, compared to electronic media.
ii) **Durability:** DNA can remain stable for up to hundreds of years, compared to up to tens of years in electronic media.
iii) **Energy efficient:** both as a function of its high data density and because DNA can be stored in room temperature without additional cooling once it is synthesized.

We therefore see that DNA storage could primarily play a crucial role in addressing the macro trend of data explosion, for use in archiving. Multiple hurdles remain for this to happen however, including the cost and speed for synthesizing and sequencing DNA, which roughly corresponds to writing/reading data. Despite that the cost has decreased a lot in the last 15 years, leading companies in the DNA synthesis space claim they need to be further reduced by a factor of a million for DNA storage to be economically viable. We do see innovation in the oligo synthesis space in the Nordics, with Karolinska Institute and KTH Royal Institute of Technology at the forefront, but these innovations are unlikely to be applied in the computing field in the near to mid term.

Neuromorphic computing (NC)

While biological components such as actual neurons can be used in BC, NC is rather a brain-inspired computing architecture with artificial neurons and synapses, which can be achieved using digital or analogue processing techniques. It differs from von Neumann architectures in multiple ways:

i) Both the neurons and synapses store information and process data, as opposed to the separation of memory and processor.
ii) All the neurons and synapses can be activated simultaneously, making it a highly parallel architecture by design.
iii) Neuromorphic chips are asynchronous, i.e., based on event-driven computation when (new) data is available.
iv) It communicates with spikes, which encode numerical data in the shape and timing of these spikes, and programs are defined by the structure of a spiking neural network (SNN).

There are two main potential benefits of a neuromorphic architectures. The first is that they are energy efficient, as they are event-based and thus only consume power when a spike occurs. Moreover, the colocation of memory and processor further improves the energy efficiency, as data does not need to be transferred between memory and storage. The second potential benefit is that they could be significantly faster than classical architectures when it comes to real-time computations. This is due to their inherent parallelism, their event-driven processing, and their low-latency.

An example of a near-term application where a neuromorphic chip would excel is processing signals such as videos. A neuromorphic chip, whose architecture is intrinsically designed to only react when there is a significant change in the data, would be optimal for this task. Changes in a frame would lead to a spike in the input, which would then be processed and stored highly parallelised across the chip, and yield a near real-time output.

Although NC addresses many of the macro trends, the area is still at an early stage with many challenges, not least on the programming and software side. We see a sparsity of startups in this field with ~30 non-stealth startups globally. [16] The most known example of a neuromorphic chip is probably Intel's Loihi, and even though truly neuromorphic ventures are a rarity, there are a few digital in-memory computing startups in Europe such as Axelera AI and Synthara.

We, however, see that the area is gaining traction, with the number of published articles and patent applications growing by 50-60% per year in the last 10 years [11,12], following an exponential development. The number of publications within NC has recently reached the levels of quantum computing publications in 2019, see Exhibit 4 in the next section. Although we are yet to see a Nordic venture within this space, and despite the many challenges for the architecture, we remain alert for Nordic research within this field and follow the research output conducted at e.g., Aarhus University, KTH, and Luleå Technical University.

Optical computing (OC)

Also known as photonic computing, technologies within OC use photons (light) and their properties to represent, transmit, and process data – as opposed to electrical signals, and binary encoding as is the case with classical computing. Technologies within photonics can be used for other types of computing, such as quantum, but we classify that as another domain. Without digging too deep into the physics behind optical computing (although we would love to), there are a few characteristics of OC that are worth mentioning:

i) Photons travel at the speed of light, resulting in exceptionally high-speed data transfer and low latency.
ii) Multiple wavelengths can be sent through the same interconnect at the same time, allowing for high bandwidth.
iii) Photons don't have mass, and therefore don't generate heat through resistance as electrons do as they move through conductors.
iv) Less energy is required to generate, manipulate, and transmit light, and generate much less heat that needs to be cooled down, making it much more energy efficient.
v) Photons in a photonic chip can be manipulated and processed simultaneously, enabling multiple computations to occur concurrently.

So, what could OC be used for? First off, we disregard fiber-optic cables, which transmit data in the form of light, and have been used for high-speed data transmission over long distances for quite some time. So, if we focus on computing as we know it, we believe that the most adjacent applications are within data transfer, primarily between chips as conversion from electronic signals to optical and back to electronic is energy consuming. Interconnect is also where we find most startups within OC, such as Black Semiconductor and Ayar Labs.

On the processing side, OC shows promises of highly parallelized computing, due to point 5 above. This parallelism makes optical processors highly effective for tasks such as large-scale AI models. But there are more challenges to overcome compared to optical interconnect, including integration of complex components onto a chip and developing signal processing algorithms, to name two. There are companies within this field, e.g. Lightmatter and Lightintelligence, and we are monitoring the processing side of OC as well.

We do not put significant focus on the memory side, simply because storing data in the form of contained photons is quite the challenge and, as far as we see it, nowhere near commercialization. Such technologies could however reduce, or even remove, the need for conversion between the electrical and photonic domain, which would accelerate the entire field tremendously. Without optical memory, we also do not foresee a full OC system anytime soon.

More ventures are emerging within the OC space, and both the number of publications and number of patents have increased by 10-20% in the last 10 years [11,12], which we hope leads to a further acceleration of the space. We observer particularly strong research within OC in multiple Nordic hubs, e.g. around Lund University, University of Copenhagen, and University of Eastern Finland - and we look forward to seeing Nordic OC startups pop-up and are particularly excited about optical interconnect innovations.

Quantum computing (QC)

Some of you might have read the 6th trend in the previous section and instinctively thought "Quantum". Many have heard about the promises of QC, and the hype around the field has exploded in recent years with over €2B invested into quantum technology startups in 2022 alone [13]. Despite the fact that vast amounts of venture capital have been invested in the field, our view remains the same as stated in our quantum thesis in 2021 – That a quantum winter is coming and that a complete QC system lies much further away than most quantum enthusiasts claim. It does not mean that we completely disregard investment opportunities in quantum technology however.

Within the broader realm of quantum technology, our investment strategy is focused on enabling quantum technology for e.g., timing, sensing, communication and networking. In that regard, we recently invested in Adamant Quanta, who employ diamond-based quantum technology for navigation and timekeeping. However, in this thesis we will focus solely on quantum computing technology.

Now, there are numerous comprehensive sources online that describes QC and it's potential, so we will describe it very briefly here. There are multiple types of quantum computers, based on superconducting qubits, trapped ion qubits, photonic quantum computers, and much more. They all differ from classical computing in many ways, some of the major ones are:

- i) Qubits are used instead of classical bits for representing information, and instead of holding a distinct state of 1 or 0, they can hold a superposition of both 1 and 0.
- ii) Multiple qubits are correlated through entanglement, as opposed to classical bits that are independent from each other.
- iii) Quantum gates are used, and they can perform fundamentally different operations than logic gates such as AND, OR, that follow Boolean logic principles.

The value of functioning quantum computing is enormous, simply put because it can address all the potential goldmines in the 6th trend, and more. When looking at the potential goldmines, there are 5 things needed to unlock the value:

i) Industry expertise to guide which specific problem that is valuable to solve significantly faster.
ii) Translation of said problem from the classical domain to the quantum realm.
iii) Middleware/control software, for e.g., error mitigation and correction
iv) Programming abstraction layers
v) Hardware to do the actual computations

So, what areas of QC are we investing in? There are already multiple QC hardware companies that have raised staggering amount of capital, but building QC hardware is a capex-intensive journey, and it is today not clear which type of quantum computers that will be superior to others, if any. We will therefore not prioritize QC hardware ventures. Programming abstraction layers are of interest but require a more mature hardware ecosystem to be really relevant, and thus we do not prioritize that either. Since industry expertise is sadly not a venture to invest in, and to our knowledge the same goes for translating problems from the classical domain to the quantum domain, we ignore those parts too.

What remains and what we first and foremost are interested in is ventures within the third buckets above, ventures focusing on middleware such as Cambridge spinout Riverlane who focus on the error correction stack and raised a ~$19M Series B earlier this year. In addition to this, we are interested in companies that provide the underlying infrastructure for the hardware field, as long as it is agnostic to the technology stack.

And if one is interested in investing in the potential of QC, the Nordics is an excellent area to look for startups within the space. Espoo in Finland is home to companies like IQM and SemiQon, while Gothenburg in Sweden hosts Chalmers University of Technology with the ~€100M research programme Wallenberg Centre for Quantum Technology (WACQT) and spinouts like Atlantic Quanta. In addition to this, there is the Quantum Computing Programme in Copenhagen, Denmark, which focus on QC for life science, with the support of ~€200M from Novo Nordisk Foundation.

We therefore look forward to seeing more ventures spinning out from multiple Nordic universities, in particular the around hubs of Espoo, Gothenburg, and Copenhagen.

Additional innovations

There are of course other research initiatives within the realm of computing hardware that are worth having a look at. One of these are Ising machines that can be used to solve combinatorial optimization problems, where spintronics is a promising field of research. Another is neuromorphic photonics, which is a fusion of OC and NC. There are also other in-memory computing architectures, i.e., memory and processor in the same unit, that work with binary input/output as opposed to spiking input/output as is the case with NC.

For current and future founders

We view this thesis as our roadmap for navigating the dynamic landscape of computing innovation, and understanding what innovations we can expect, their potential, and their challenges. Industrifonden is a venture fund with Nordic mandate and can therefore not be niched to focus on one specific sub-field of one specific architecture. We observe that the entire field of computing is gaining attention, with the number of publications and patents increasing rapidly as demonstrated in Exhibit 4.

Research in NC, OC and QC has increased rapidly; NC publications are at the same level as QC was 3 years ago

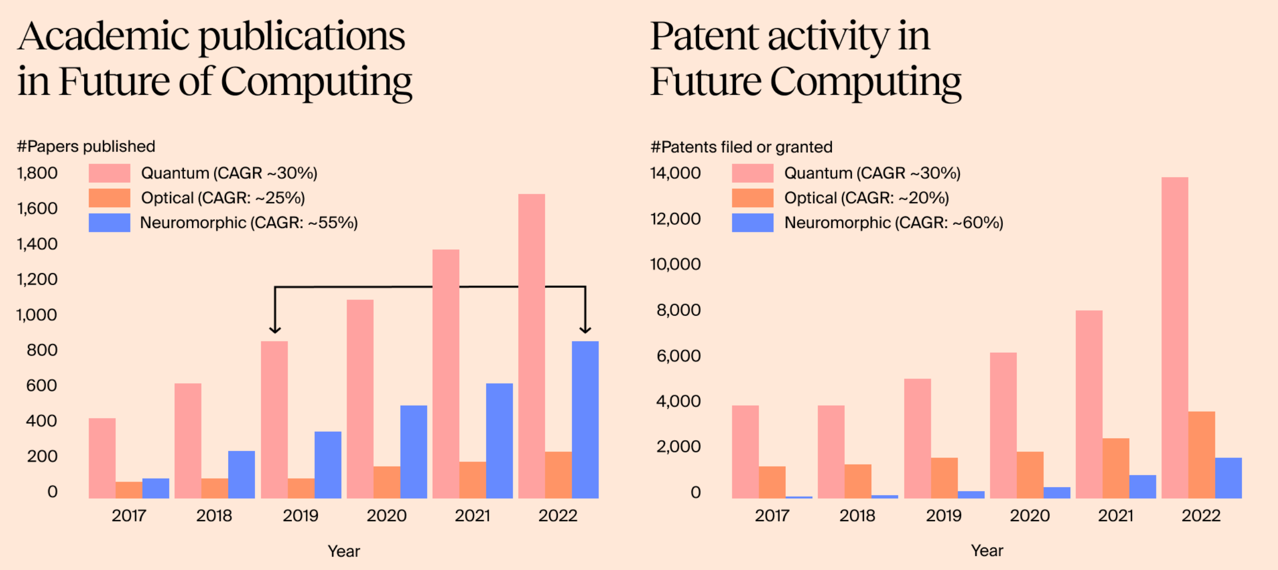

Exhibit 4. Sources: Patsnap [11]; Web of Science [12]

Our interest thus lies in the entirety of innovation within the realm of the Future of Computing, and we hope that this thesis can help inspire excellent scientists within the broad realm of computing to commercialize their innovations so that they can make a long-lasting impact on society.

While we refrain from limiting our focus to a specific sub-field or architecture, we do acknowledge that certain areas hold significant promise in terms of their potential impact and likelihood of success. These four areas, in particular, have captured our attention:

- i) Accelerated computing innovation that focus on memory and interconnect in existing architectures
- ii) Optical computing innovation for interconnect within and between chips
- iii) Middleware for quantum computing
- iv) In-memory compute, e.g. neuromorphic chips

Future of computing is a major deep tech focus for Industrifonden, and we look forward to investing more within the computing space. After all, it is not a coincidence that 3 of the 6 meeting rooms at Industrifonden's offices are named after some of the computing pioneers mentioned in the Introduction.

References

1. **McKinsey & Company.** The state of AI in 2022—and a half decade in review. [Online] December 6, 2022. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

2. **Stanford HAI**. AI Index Report 2023.

3. **Sevilla et al. 2022**. *Compute trends across three eras of machine learning*. arXiv:2202.05924v2

4. **Ericsson**. Ericsson Mobility Report 2023.

5. **Gartner**. Emerging Tech Roundup: Navigating Confusion Around Synthetic Data. 2023.

6. **Statista.** Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025. [Online] August 22, 2023. https://www.statista.com/statistics/871513/worldwide-data-created/

7. **Intel.** Intel Launches 4th Gen Xeon Scalable Processors, Max Series CPUs [Online] January 10, 2023. https://www.intel.com/content/www/us/en/newsroom/news/4th-gen-xeon-scalable-processors-max-series-cpus-gpus.html#gs.5wv3j0

8. **NVIDIA.** NVIDIA H100 Tensor Core GPU. [Online] September 20, 2023. https://www.nvidia.com/en-us/data-center/h100/

9. Nicola Jones, 2018. "How to stop data centres from gobbling up the world's electricity," **Nature**, vol. 561(7722), 163-166.

10. **Capgemeni.** Gunnar Menzel, "*The (more) Sustainable Data Center*" [Online] January 28 2020. https://www.capgemini.com/insights/expert-perspectives/the-more-sustainable-data-center/

11. **Patsnap.** Data retrieved August 2023.

12. **Web of Science.** Data retrieved August 2023.

13. **The Quantum Insider**. Quantum Technology Investment Update 2022 Review.

14. **Gartner analyses.** Data retrieved September 2023.

15. **BCG**. Jean-François Bobier, Matt Langione, Edward Tao, and Antoine Gourevitch. What Happens When 'If' Turns to 'When' in Quantum Computing? July 2021

16. **Pitchbook.** Data retrieved August 2023.