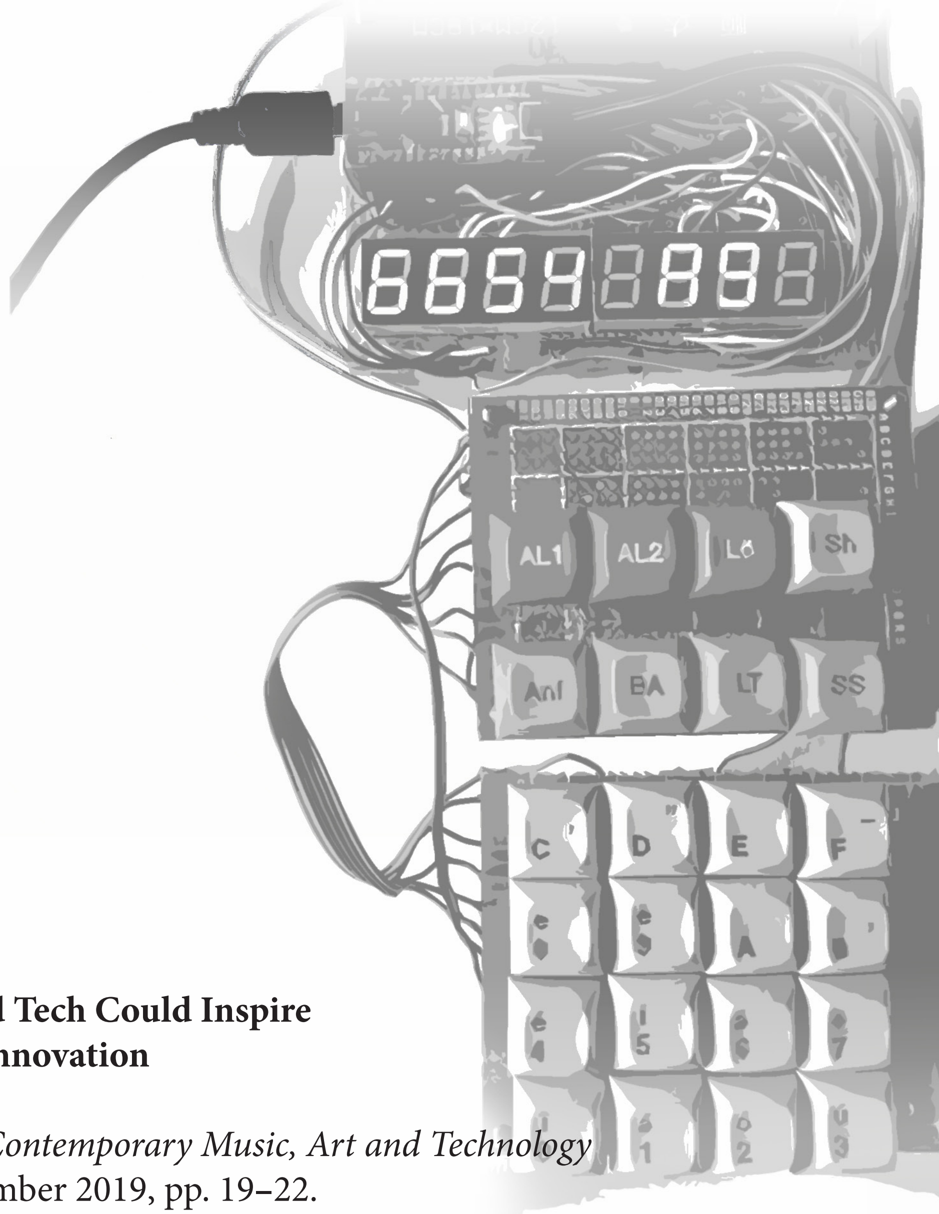
I N S Δ M

JOURNAL OF CONTEMPORARY MUSIC, ART AND TECHNOLOGY



How a Century-old Tech Could Inspire the Next Wave of Innovation

Adnan Mehonić

INSAM Journal of Contemporary Music, Art and Technology
No. 3, Vol. II, December 2019, pp. 19–22.

INSΔM

Adnan Mehonić
Department of Electronic and Electrical Engineering,
University College London
London, United Kingdom

HOW A CENTURY-OLD TECH COULD INSPIRE THE NEXT WAVE OF INNOVATION

Lee Sedol, 18-time Go world champion, has recently announced his retirement from the professional play. What makes the announcement worth thinking about is the reason for walking away from the endeavour Lee dedicated his life to - "Even if I become the number one, there is an entity that cannot be defeated."[1]

In 2016, Google's DeepMind artificial intelligence system (AlphaGo) had beaten Lee in Go (4 to 1) – an ancient game that represented a Holy Grail for AI for a long time. The reason why researchers have adopted Go as a good benchmark for intelligence is the fact that a brute-force strategy to play the game does not work, but a level of "creativity" is needed. A back-of-the-envelope-calculation estimates 10^{800} possible matches of a typical length, which outnumbers the total number of atoms in the known universe by the factor of 10^{720}. Taking this into account, we cannot be blamed to consider AlphaGo having a particular type of intelligence, and we are intrigued by what might come next (well beyond the newer generation AlphaGo Zero, that beat the previous version by the score of 100 to 0). Self-driving cars, intelligent self-diagnostics

1 https://en.yna.co.kr/view/AEN20191127004800315 (on 12/12/19)

tools, autonomous robots, knowledgeable personal assistants, and more powerful tools for surveillance are some of the promises of AI, and not to forget that AI is now producing what some consider as a form of art.

However, here is the other side of the coin. There is no such thing as a free lunch (at least not in this case). AI is not magic that happens without costing something: on the contrary, it might cost us more than we think. When we speak to Siri or any smart personal assistant of choice, it costs us a certain (typically vast) number of computations, and that has its price. To produce AI, we need to train artificial neural networks, and this process is usually run on large computer clusters. We dubbed this process as cloud computing, but the service does not happen somewhere in the atmosphere but in immense and extremely power-hungry data centres. Training neural networks requires a lot of energy, and that has its consequences. Recent estimates suggest that training a few neural networks could produce as much as 284 tonnes of carbon dioxide.[2] To put things in perspective, this is equivalent to the lifetime emission of 5 average cars. Even more alarmingly, it has been suggested that data centres could become one of the biggest polluters in the world, consuming one-fifth of the total energy produced on Earth as early as 2025.

The trend does not seem to be slowing down, and if anything, it is increasing at a fast pace. Since 2012, the computing power requirements to keep AI going have increased by a factor of 300,000, remarkably more than what would be expected by historical trends (closer to an x7 increase).

So, it makes sense to think from first principles and see if we can do better.

Lee Sedol (+coffee) taking on DeepMind's AlphaGo (March 2016)
(Image adapted from: DeepMind/YouTube)

2 E. Strubell et al, Energy and Policy Considerations for Deep Learning in NLP. arXiv:1906.02243v1 (2019).

In his 1965 paper, Gordon Moore, co-founder of the Intel Corporation, predicted that the number of transistors, which are fundamental building blocks of electronics, in integrated circuit doubles every 18–24 months. This prediction is better known as Moore's Law. The ability of the semiconductor industry to follow this trend through transistor scaling provides a better performance-to-cost ratio of products and results in the exponential growth of the semiconductor market. What every transistor individually does is simple. It switches between two states: 0 and 1. However, where the complexity emerges is when we combine billions of transistors into a single electronic device; the Apple iPhone's A13 chip has 8.5 billion transistors, and it does some impressive things. Moreover, Dennard's Law, a close cousin of Moore's Law, observes that the transistors are also getting more energy-efficient and faster as we scale them. It seems the only (main) thing we had to do to fuel our electronic technology was to keep making transistors smaller and smaller, and as a lucky side-effect, we got them to be more energy-efficient as well. This is a digital approach, the basis of virtually all electronic devices we use today, and the paradigm has been so successful for the last five decades that we have taken it for granted. However, all of this is about the change. The trend of making transistors smaller has significantly slowed down in the last few years, marking the beginning of the end of Moore's law. The trend of them becoming more efficient has practically stopped ten years ago. The issue is that we cannot go beyond physical limits - it is challenging to make transistors consist of a handful of atoms that make all the matter and still having them to be fully functional. The electronics cannot work at faster speeds as they will melt due to the extensive heating. Simply put, we cannot beat physics. So what can be done? It seems we need a new approach, or maybe we can look at the old obsolete technology – analogue computing.

Analogue computing is a long-forgotten endeavour, being abandoned almost half a century ago, mostly due to the enormous success we had with a digital approach. Analogue components are prone to noise, easily affected by environmental changes, and less reliable in producing entirely predictable results. After all, a calculator that makes an arithmetic mistake even once is not good enough. However, analogue systems have one crucial advantage: they are governed by the same equations they try to solve, leading to much better energy efficiency. Instead of using millions of digital transistors to simulate and solve a problem, we can use the physics of a handful of analogue components more directly to get a result. If only we could somehow deal with the imprecisions of analogue components. Well, we do know of one system that is not digital, is composed of messy and noisy elements, but still produces the wanted results with remarkable energy efficiency. Here comes the human brain, the most complex, intelligent, but also energy-efficient system we know of. Neurons and synapses are intrinsically noisy, messy, and stochastic, and yet the brain functions and produces intelligence better than any other system we know. Putting things in perspective, AlphaGo consumes around 1MW of power and uses all of those watts to compete against Lee's brain, which does much more than play a game and still does not consumes more than 20W. Maybe we can learn something from biology and try to resurrect some of the old analogue approaches. Only, this time, we also have all the benefits of novel technological developments, such as those in nanotechnology.

I work on such a novel technology called memristors.

Memristors are nanoelectronic devices that can be as small as our current smallest transistors, but that is not their ultimate strength. Memristors aim to be much more than simple digital switches. They enable computing by directly implementing some crucial functions of biological systems—most importantly, synapse-like plasticity and neuron-like spiking. Memristors are much closer to analogue components than digital transistor switches. They are not perfectly deterministic, and they do exhibit some level of stochasticity. However, the whole paradigm of artificial neural networks, our current bedrock of AI, is to harness systems that use probabilities (the AI system rarely predicts answers with 100% confidence). Memristors, when utilised in physical neural networks, still provide excellent accuracy while using a small portion of the energy budget. There is a chance that memristors will be a key enabler for low-power AI systems of the future.

In either case, it is sometimes worth combining new with the old. For a researcher, this might mean going to a library and digging into the yellow pages of old, non-digitalised manuscripts.

A "calculator" is better for staying digital; however, analogue might have its place in the future of AI, and they could both coexist. By combining old ideas with new technologies, we might be on the right track to produce the next paradigm that might keep us running for the next few years, if not decades, as in the case of Moore's law.