Considerations on transformative AI and explosive growth from a semiconductor-industry perspective

Muireall Prase (muireall.prase@proton.me)

May 30, 2023

Summary

This essay attempts to bridge a gap between abstract models of AGI timelines and inside views from the semiconductor industry.

Models like those in Cotra 2020 (“Bio Anchors”), Davidson 2021a (“Explosive Growth”), and Davidson 2023 (“Compute-Centric Framework”) are grounded in computational and economic abstractions. This is entirely appropriate for medium-to-long-range forecasting, where extrapolating from trends at large is generally more reliable than reasoning about processes in detail.

Even so, I think their cursory accounts of some of their most important parameters are major weaknesses. A complementary approach works outward from an inside view—not by immediately pursuing better parameter estimates, but by first trying to uncover as many relevant considerations as possible. To the extent these considerations make contact with our models’ abstractions, we can use them for model inputs; anywhere they don’t, we can rethink our models.

The body of this essay is a short scenario-planning exercise meant to demonstrate this approach, considering two pathways by which AI might fail to have transformative impact in the coming decades. The scenarios are chosen to invite considerations from the semiconductor industry—both its history as a point of comparison and its future as part of the cycle of AI progress. The industry is as mature as any (perhaps second to parts of the chemical industry), notably in terms of production scale and optimization up against physical limits. It’s grown faster than the global economy for

decades, feeding improved hardware back into itself but without explosive (superexponential) growth. AI aspires to the same scale but is comparatively immature. It leans more heavily on software, which has very different margins and gets less durable competitive advantage from intellectual property. It may also enter a similar feedback cycle, although there's a significant potential difference in how effectively AI outputs can substitute for labor.[1]

[1] That's the difference at the core of how advanced AI could drive explosive growth in Davidson 2021a and Davidson 2023.

I use two scenarios to capture different kinds of obstacles and bottlenecks to AI progress. In the first, firms struggle to capture returns on investment in AI R&D. In the second, AI progress fails to substantially accelerate the non-AI inputs to AI R&D. In both scenarios, highly capable AI may be achieved, but it does not arrive suddenly or have transformative impact on short timescales.

This exercise, rather than try to estimate parameters for economic modeling from abstract considerations like the above, attempts to distill the two scenarios into lower-level causal drivers. I also try to identify observable indicators for the degree of influence those drivers may have. Because a comprehensive evaluation of data on these indicators would vastly expand the scope of the essay, I mainly call out examples to clarify meaning or relevance, not to argue the weight of evidence in one direction or another. Appendix A describes the process used to develop scenarios and other elements of the analysis in more detail.

Other appendices provide context and support for some claims that are non-obvious but incidental to the scenario analysis. Appendix B briefly discusses the pipeline from research to deployment in semiconductor device production. Appendix C is an extended discussion of problems with the nanomechanical computer described in Eric Drexler's *Nanosystems*. Finally, Appendix D contains some related forecasts to more transparently convey my own expectations.

Ideally, this kind of exercise would be done by a panel with overlapping areas of expertise. My effort is far from exhaustive, but I hope it at least shows how this approach might usefully fit into a broader forecasting or planning project. If I wanted to single out a few particular themes I think existing analyses don't fully appreciate:

1. Models of transformative change would strongly benefit from deeper inside views.
2. Hardware specialization is a meaningful obstacle to rapid growth through paradigm shifts or redeployment of resources.

3. A robust open-source community and publicly-owned AI services can dampen AI growth along paths that require large, lumpy capital investment, in part by reducing the potential for AI ventures to capture profits.

4. Progress in industry at scale is not limited by "ideas" in the same way that basic research could imaginably be.

Scenario 1

In this scenario, profits never materialize and venture capital thins out. Nonprofits and governments are still willing to fund projects for public benefit or military advantage, but there's no strictly financial virtuous cycle in which returns on investment accrue to investors, bypassing the economy as a whole.

As far as there are periods of rapid technical progress, they are self-limiting. AI progress is "lumpy", each step requiring large initial investments and long training runs on a fixed course. As long as the quantity of inputs you can buy at a given price is growing much faster than the rest of the economy, it's often better to wait to train a better model at the same cost. Uncertainty about the landscape by the end of a training run also moderates investment. [2] Explosive growth would entail unprecedented "creative destruction", as noted in Jones 2021. Competition keeps margins low, and obsolescence is a sharp cliff: any given product has only until a better model is trained to recoup training costs.

Research nevertheless continues deep into the current paradigm. Further incremental advances face diminishing returns as far as they stay within that paradigm. Breakthroughs departing from it, on the other hand, have to compete with accumulated years of research and investment: specialized datacenters of specialized hardware, mature software architectures and data pipelines, and a large pool of research and engineering expertise. The expected benefit of bringing a potential breakthrough to scale is rarely worth any catch-up costs that could instead be invested in steady progress over the same period.[3]

This is analogous to the situation the semiconductor industry finds itself in. For a time, pre-competitive collaboration through industry roadmaps and consortia like SEMATECH allowed players to share development costs towards common targets 10–15 years in the future. Semiconductor consortia still exist, but simultaneous industry fragmentation and leading-edge

[2] Often attributed to Robert Palmer: "Designing microprocessors is like playing Russian roulette. You put a gun to your head, pull the trigger, and find out four years later if you blew your brains out." More abstractly, this is the sort of equilibrating dynamic that can help maintain the "knife-edge" condition of constant exponential growth considered in Davidson 2021a.

[3] This dynamic is illustrated repeatedly in Cyrus Mody's *The Long Arm of Moore's Law: Microelectronics and American Science*: "[Richard Smalley's startup Carbon Nanotechnologies, Inc.]'s nanotubes never made their way into IBM's commercial transistors, or anyone's, in large part for the same reason that Josephson computing and the earlier generations of molecular electronics didn't succeed. Silicon moved right past all its competitors, as predicted by Moore's Law." (Mody 2016, p. 211)

foundry consolidation has reduced their membership and influence today. The AI industry finds itself unable to follow this example. Short-term unpredictability and competitive pressures limit the size and length of bets industry players are willing to make.

In this scenario, AI will contribute to social and economic change, but it is not on a path to reach transformative capabilities.

Key drivers and indicators

Key drivers towards this scenario relate to profitability of AI research and production:

1A Specialization of hardware and infrastructure for a particular paradigm of AI

1B AI training runs as major capital projects

1C Difficulty capturing value from training large AI models

1D Uncertainty about returns from scaling new methods

Indicators—hypothetical observations that would lead us to expect the future to have more in common with this scenario:

1a. Technologies like TPUs, low-precision arithmetic, and specialized servers and data centers dominate general purpose CPUs, consumer GPUs, and traditional datacenters for use in AI.[4] (+A)

1b. Training and deployment in particular use different hardware and infrastructure. A datacenter used for training can't be turned around and used for inference without suffering major inefficiencies. Hardware specialized for LLMs puts new paradigms at a disadvantage out of the gate. (+A, +B)

1c. Generic progress in the semiconductor industry only marginally advances AI hardware; for example, AI hardware lags general-purpose hardware in switching to new process nodes. (+A)

1d. Conversely, advances in AI hardware are difficult to repurpose for the rest of the semiconductor industry. For example, AI training requirements drive High Bandwidth Memory development, but GDDR or LPDDR suffices for other high-volume, high-performance applications.[5] (+A)

[4] For example, we observe that Meta has canceled or paused datacenters mid-development and has recently described its next-gen datacenter design in a pivot to generative AI workloads.

[5] Some discussion of the differences here.

1e. Specialized hardware production is always scaling to meet demand; at any given time, essentially all suitable hardware is deployed in AI. Proxy indicators include the growth of specialized firms. (+A)

1f. Research progress is driven chiefly by what we learn from the largest and most expensive projects. Large models can't be effectively understood using small proxies with the same architecture. New systems have unpredictable capabilities, outside expectations from "scaling laws"; progress is even non-monotonic in some domains. (+B, +D)

1g. Open-source models and second-tier competitors lag the state of the art by around one large training run. (+C, +D)

1h. Small models can be cheaply trained by using expensive models for evaluation, achieving results nearly as good at much lower cost. Model weights are leaked, pirated, or otherwise exfiltrated. (+C)

1i. Progress in capabilities at the frontier originates from small-scale experiments or theoretical developments several years prior, brought to scale at some expense and risk of failure. (This is the status quo in hardware.) (+D)

Negative indicators—hypothetical observations that would lead us to expect the future to have less in common with this scenario:

1j. The same hardware pushes the frontier of performance per watt not only for AI training and inference but also for high-performance computing more traditionally. (–A)

1k. Emerging hardware technologies like exotic materials for neuromorphic computing successfully attach themselves as adjuncts to general-purpose silicon processes, giving themselves a self-sustaining route to scale. (–A)

1l. Training runs use as much compute as they can afford; there's always a marginal stock of hardware that can be repurposed for AI as soon as AI applications become slightly more economical. (–A)

1m. Industry players engage in pre-competitive collaboration, for example setting interoperability standards or jointly funding the training of a shared foundation model. (–B)

1n. Alternatively, early industry leaders establish monopolistic advantages over the rest of the field. (–B, –C)

1o. AI training becomes more continuous, rather than something one "pulls the trigger" on. Models see large benefits from "online" training as they're being used, as compared with progress from model to model. (–B)

1p. Old models have staying power, perhaps being cheaper to run or tailored to niche applications. (–C)

1q. Advances in AI at scale originate from experiments or theory with relatively little trouble applying them at scale within a few years. (This is the status quo in software.) (–D)

1r. The leading edge features different AI paradigms or significant churn between methods. (–A, –D)

Scenario 2

In this scenario, the balance tips and investment pours in. AI technology advances, new markets are unlocked, more money comes in, more research is done. This is not an exceptional pattern. Government subsidies kickstarted the virtuous cycle in renewable energy technology,[6] as did defense spending in semiconductor research. Sometimes the cycle involves less R&D and looks like simple ,"returns to scale" ("learning rate", "experience curve")—every time cumulative production doubles, per-unit costs decrease by some fraction. Nothing about this necessarily involves direct feedback from outputs to inputs.

[6] Roser 2020.

If more advanced AI makes further advances easier, we might expect the AI industry's learning rate to grow over time, leading to explosive growth. Still, AI here isn't obviously qualitatively different from the semiconductor industry, where continued progress leans heavily on modern computers, or even from the energy industry, where cheaper power makes everything easier. In the medium term, such positive feedback balances at some level with diminishing returns—both "intrinsic" (the same spending yields less performance progress) and "extrinsic" (as one input to production becomes less costly, it becomes a smaller fraction of total production costs, so further improvements will be less beneficial—as in Amdahl's law).

In this scenario, growth is not explosive, but instead follows roughly the same curve it's been following. AI becomes capable of assisting in AI research and development, perhaps even unpredictably and rapidly surpassing human efforts, but this looks more like a one-time decrease in costs per “unit” of R&D. Other inputs like hardware and logistics do not keep up in efficiency gains, so such inputs soon dominate the pace and price of both AI R&D and practical deployment.

One important bottleneck is science. Experiments take time and aren’t particularly bound by human schedules or a shortage of “ideas” to begin with. Individual steps have non-trivial physical rate limits: temperature ramp times, silicon pull rates, vacuum pump speed, chemical reaction rates, measurement bandwidth. Many experiments need to happen serially. Where it matters, experimental design is already highly optimized by humans. AI is most helpful when it can replace experiments with simulation, but physical simulation is intrinsically hard, already highly optimized, and rarely exact enough to obviate experiments. Neural networks as surrogate models accelerate many results, but sometimes we still have to solve the equations the way we already do.

Even within software, AI R&D isn’t bottlenecked by how fast code or ideas are generated. Comparing increasingly highly capable AI systems is subtle, and so requires time-consuming computational experiments to evaluate any changes. Models are not fully general or are assembled from interacting specialized modules that can’t be evaluated on their own, so it takes many separate experiments to advance the frontier.

Finally, it turns out there’s just not that much room for improvement. Gains from hardware top out at two or three orders of magnitude from the remnants of scaling and architecture specialization. Exact algorithms are already close to optimal, and making faster approximations doesn’t yield superintelligence. Photonic computing is limited to niche applications outside high-performance computing. Quantum advantage is minimal outside a narrow set of problems. Reversible computers are impractical. Nanomechanical computers are unpromising (Appendix C).

As a result, AI growth remains tethered to the rest of the economy. AI may drive economic growth, but transformative potential is bottlenecked by activity other than that of deployed AI.

Key drivers and indicators

Key drivers towards this scenario relate to how much AI can accelerate the non-AI inputs to AI R&D:

2A Limited generality of AI applicability

2B Limits on headroom for progress

2C Bottlenecks in logistics, construction, and other serial physical activity

2D Difficulty of substituting theory or simulation for experiment

Indicators—hypothetical observations that would lead us to expect the future to have more in common with this scenario:

2a. Progress in AI is very uneven across domains—each faces different bottlenecks that are addressed individually. (+A)

2b. Apparent technical wins—upgrades with predictable effectiveness—are left on the table, because they only affect a fraction of performance and impose adoption costs on the entire system. (+B, +C)[7]

2c. The semiconductor industry continues to fragment. Large consortia die out. Progress comes not from coordinated industry-wide innovation but from specialization in all possible directions.[8] (+B)

2d. More broadly, semiconductor industry trends continue: foundry consolidation continues, foundry construction costs and times continue to rise, chip design costs continue to rise, fabrication costs continue to rise, all exponentially and with diminishing returns. (+B, +C)

2e. Semiconductor industry roadmaps continue to extend 10–15 years out. The roadmaps are stable, suggesting that "ideas" are not a bottleneck. Gate-all-around (GAA) field effect transistors (FET) replace FinFETs in new process nodes; Complementary FET (CFET) designs succeed those, and they do so no sooner than 2032. (+C, +D)

2f. AI research progress is driven by training large models and seeing what happens. (+D)

[7] This is a bit oblique, but it's an indicator of how "bottlenecky" a domain is. One example might be what happened to 450 mm wafers. Even though you get 2.25x more area per wafer (over the current 300 mm standard), a large enough fraction of costs scale with area such that the net cost per area may only go down by 20–25%. At that point, it becomes difficult to justify the cost of developing new equipment and upgrading fabs.

[8] Leiserson et al. 2020.

Negative indicators—hypothetical observations that would lead us to expect the future to have less in common with this scenario:[9]

2g. The same general AI is broadly deployed in different domains, industry coordination is strong (through monopoly or standardization), and upgrades hit many domains together. (–A)

2h. Evidence builds that a beyond-silicon computing paradigm could deliver performance beyond the roadmap for the next 15 years of silicon. (See Appendix C for an example of what I would not count as evidence.) (–B)

[9] The positive indicators are basically the status quo, and these are basically "something revolutionary happens". With more time and imagination one might be able to write something more useful.

2i. New semiconductor consortia arise, for example producing consensus chiplet or heterogeneous integration standards, making it easier for a fragmented industry to continue to build on one another's work. (–C)

2j. Spatial/robotics problems in particular—proprioception, navigation, manipulation—are solved. (–C)

2k. Fusion power becomes practical. (–C)

2l. AI is applied to experimental design and yields markedly better results than modern methods. (–D)

2m. AI research progress is driven by theory. (–D)

2n. Breakthroughs make microscopic physical simulation orders of magnitude easier. Molecular dynamics, density functional theory, quantum simulation, and other foundational methods are accelerated by AI while also greatly improving accuracy. (–D)

Appendices

Appendix A Selecting methods, scenarios, key drivers, and indicators

I am a scientist working on experimental special-purpose electronic hardware unrelated to machine learning. I have no affiliation with any organization mentioned in this document. What I've written here does not represent the views of my employer.

The seed of this document is a sense that something is missing in existing writing about catastrophic risk from AI, arising in particular in reaction to Cotra 2020 ("Bio Anchors"), Carlsmith 2021 ("Power-Seeking AI"), Davidson 2021a ("Explosive Growth"), and Davidson 2023 ("Compute-Centric Framework").[10] The level of discussion is generally abstract, even in situations where detail is relatively easily available and informative.

From "Bio Anchors", on estimates of growth of effective FLOP per dollar, historically, in the medium term, and in the long term: "Because they have not been the primary focus of my research, I consider these estimates unusually unstable, and expect that talking to a hardware expert could easily change my mind." On room for medium-term improvements in silicon,[11]

> Of all the quantitative estimates in this document, I consider these forecasts the most likely to be knowably mistaken. While most of the other quantitative estimates in this document have a lot more absolute uncertainty associated with them, there is a lot more low-hanging fruit left in improving short- and medium-term hardware price forecasts. For example, my understanding is that semiconductor industry professionals regularly write highly detailed technical reports forecasting a number of hardware cost-efficiency metrics, and I have neither read any of this literature nor interviewed any hardware experts on this question. Paul is a machine learning researcher, not a hardware expert, and he only spent a few hours thinking about this question; I spent even less time discussing it with him.

Or consider "Compute-Centric Framework" on what evidence we have about substitutability for AI inputs:[12]

> The third bucket of 'evidence' is simply doing inside-view thought experiments[13] about what you think would happen in a world

[10] Others include Carlsmith 2020; Davidson 2021b, and several related blog posts on Cold Takes, Open Philanthropy, and Less Wrong. I'm long familiar with arguments about AI as an existential risk, although I've only recently begun to catch up on writing about AI alignment from the past several years.

[11] I'm not pointing this out as a critique of "Bio Anchors"—Cotra correctly notes that variation within the range of uncertainty here does not affect its conclusions. At the same time, becoming much more familiar with these sorts of details can lead one towards qualitatively different approaches to thinking about the future.

[12] Davidson 2023, Part 2, p. 15.

[13] I would not call what is described here "inside view", except perhaps in a relative sense.

> with zillions of AGIs working on (e.g.) hardware R&D. How much more quickly could they improve chip designs than we are currently, despite having access to the same fixed supply of physical machinery to use for experiments? If you think that hardware progress would be 100X its current pace, you can use this to "back out" a value of ρ [substitution parameter in the CES production function] consistent with that. This type of thought experiment gets at ρ for cognitive tasks vs non-cognitive tasks....
>
> Doing this kind of inside-view thought experiment gets into lots of tricky issues like "Could you replace physical experiments with simulations?" and "How many experiments would be needed for a smart enough team of AIs to discover nanotech and use it to design better chips?". These questions are, I think, worthy of much more investigation. It would be useful to think through specific candidate bottlenecks concretely and assess how much they would slow down progress.
>
> This third bucket of 'evidence' leads me, at least in the case of hardware R&D, to higher estimates of than the first bucket [empirical macroeconomic research]. If $\rho = -0.5$ and [share parameter for substitutable tasks] $\alpha = 0.3$ (as I suggested for hardware R&D), then even zillions of AGIs would only increase the pace of hardware progress by $\sim$10X. But with billions of AGIs thinking 1000X as fast and optimising every experiment, I think progress could be at least 20X quicker than today, plausibly 100X. If $\alpha = 0.3$, a 100X speed up implies $\rho = -0.25$. I expect some people to favour larger numbers still. Very large numbers would favour choosing a value of ρ very close to 0 (but still negative), which would approximate Cobb Douglas ($\rho = 0$).

This is one of the most important parameters in the model,[14] and the thought experiments or intuitions used to produce the quoted multiples are not even described, let alone spelled out. From my perspective, it's hard to imagine this happening for anything that requires new process recipes, let alone tool upgrades. That leaves improving chip architecture given fixed elements, and there may not even be enough room for improvement there for 100X faster R&D to be meaningfully measured. If we stopped building fabs now, I imagine we'd squeeze everything we could out of current processes well within 10 years. I could be convinced that zillions of AGIs could do the same in months, if only because it presupposes enough compute

[14] Well, maybe. See a bit further below.

to brute-force optimization. But I think the interpretation of CES parameters we want should exclude that presupposition from consideration—we're supposed to be pumping intuitions for how much AI can help with things like simulation in the first place. I'm not sure how to make this picture coherent.[15]

One thing I'm getting at in particular is that it's difficult to "sanity check" surprising results without some breadcrumbs back to concrete mechanisms. For example, tinkering on the "Compute-Centric Framework" model playground, it's clear that the main conclusions are robust in most parameters. Even parameters that seem like they should be important barely move the conclusions. There's also some outright counter-intuitive behavior—for example, changing various parameters in a conservative direction brings "pre wake-up", "wake-up", and "mid rampup" closer to the present. (Most simply, I've found that starting with the "best guess" preset, increasing hardware adoption delay does this; so does increasing both AGI training requirements and effective FLOP gap, keeping the ratio constant so that the FLOP requirement for 20% automation should be constant.) I worry that the apparent robustness is a sign that the conclusions about takeoff speeds are "baked in" to the model in some way—that a much simpler picture could give the same result, and the complexity of the model only serves to obscure that result.[16]

I also agree with Ben Jones in his review of "Explosive Growth":[17]

> Ultimately, I think bottlenecks are where the action is. An interesting descriptive exercise would be to consider (a) the current array of goods and services that humans consume to see which ones seem both essential to the standard of living and least amenable to a scalable automation solution and (b) the array of activities in scientific and technological advance that are amenable to a scalable intelligence. It doesn't take much in the way of bottlenecks to severely undermine the growth implications of AI, even if AI is really fantastic at very many things.

From my experience and reading about forecasting and structured analysis, some form of "scenario analysis" ("alternative futures", "scenario planning", ...) is suitable for filling in these sorts of missing causal pieces. For example, from "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis", p. 34:

> Alternative futures analysis (often referred to as "scenarios") is

[15] The actual model in question ignores physical simulation. I think footnote 38 basically agrees about the problem, but I don't know what to make of the proposed workaround: "This leaves it ambiguous whether the AGIs have access to unlimited compute for running simulations. In the FTM, the potential importance of simulations for hardware R&D is not modelled. In practice, then around and shortly after AGI there will be a fair bit of computation available for simulations like this. (They'll be at least the compute needed to train AGI, and there will be strong incentives to use compute for hardware R&D.) But there won't be ∼infinite computation sitting around. So probably the best thing to imagine for this thought experiment, is 'you have a decent amount of compute for simulations, but not an insane amount so you can't just brute force things'." Davidson 2023, Part 2, p. 21

[16] I have some similar feelings about "Bio Anchors"—for example, the convolution of 2020-equivalent FLOPs and algorithmic progress under different anchors—but in the end, it's open about multiplying together very uncertain estimates of a few key numbers.

[17] Jones 2021.

> most useful when a situation is viewed as too complex or the outcomes as too uncertain to trust a single outcome assessment. First, analysts must recognize that there is high uncertainty surrounding the topic in question. Second, they, and often their customers, recognize that they need to consider a wide range of factors that might bear on the question. And third, they are prepared to explore a range of outcomes and are not wedded to any preconceived result....
>
> Alternative futures analysis is extremely useful in highly ambiguous situations, when analysts confront not only a lot of "known unknowns" but also "unknown unknowns." What this means is that analysts recognize that there are factors, forces, and dynamics among key actors that are difficult to identify without the use of some structured technique that can model how they would interact or behave.

I'm largely deferring to the judgment of others here; I'm not aware of quantitative or systematic comparisons of different techniques like this. I also don't exactly follow any one methodology, since these generally involve multiple people, slightly different goals, and much more time than appropriate for this context. I do try to follow some recommended practices, including from an assessment of structured analytic techniques (SATs) by RAND:

> Our pilot study did suggest several best practices, particularly in documents containing alternative scenarios, the SAT employed most frequently in our sample. One area for improvement in some of the documents was, in our opinion, greater transparency about the reasoning behind the scenarios:
>
> - In particular, when intelligence analyses posit key drivers, it would be useful to know how they were selected and why they were considered more important than other potential drivers.
> - In addition, it would be helpful to explain—perhaps in a box or appendix—the methodology that was used to construct the scenarios. These steps would help maximize the value of SATs in making clear to readers how analysts reached their key judgments.

- Scenario papers also are most useful when they include concrete and observable indicators that signal which outcome is becoming more likely, rather than broad generalities that are difficult to measure.

I don't feel I've fully succeeded in the last point; all the indicators are still fairly broad generalities, owing to time constraints and the consequential emphasis on communicating a way of thinking about the problem rather than on particular conclusions. Still, in accordance with the other recommendations, I'll try to explain some of my decisions below.

Scenarios aren't entirely new to AI forecasting,[18] although they do seem underused. There is more of a gap in descriptions of "safe" scenarios. Grace 2022 is a strong exception, but given the different framing and emphasis, my approach is not redundant. I decided to focus on scenarios with AI capable of contributing to scientific research where nonetheless there is no explosive growth in economic activity or AI capability. Such scenarios help fill that gap and highlight some details that feed into my disagreements with existing forecasts on the impact of AI, many of which are rooted in my relative familiarity with semiconductor research and development.

[18] A good example for catastrophic scenarios is Sotala 2018.

Some of these details I first identified included specific "Baumol tasks",[19] practical drivers of diminishing R&D returns unrelated to ideas getting harder to find, how long things take in the semiconductor industry, and more pessimistic limits to physical and learning capabilities.

[19] Aghion, Jones, and Jones 2017.

I began categorizing these details into themes that I thought might anchor scenarios:

1. Economic and scaling considerations limit whether and how fast "breakthroughs"[20] make it to production in large-scale industries
2. Real-world logistic activity limits the pace of feedback cycles
3. Serial experimental requirements limit the pace of feedback cycles
4. Technological capabilities are generally bounded

[20] say, emerging technologies that are objectively superior in some particular sense

The best way I could find to balance these was with two scenarios, where Scenario 1 mainly captures the first point, while Scenario 2 captures the next two together. While their respective drivers are not mutually exclusive, Scenario 2 still works as conditional on avoiding Scenario 1. The last point is a vast topic and I address only a small part of it in Appendix C.[21] I attempted to describe observable indicators for relevant details, and then distilled key drivers corresponding to groups of those indicators on a slightly

[21] Interesting related work includes Markov 2014 and research by Neil Thompson and collaborators.

higher level of abstraction. From there, the scenarios, drivers, and indicators evolved together, mainly as I made the indicators more or less concrete.

Appendix B From research demonstration to production at scale

How long does it take to bring a new technology to semiconductor production? I've often heard estimates of 10–15 year lead times. For example, Joel Hruska, "How Are Process Nodes Defined"?:

> Semiconductor manufacturing involves tremendous capital expenditure and a great deal of long-term research. The average length of time between when a new technological approach is introduced in a paper and when it hits widescale commercial manufacturing is on the order of 10–15 years. Decades ago, the semiconductor industry recognized that it would be to everyone's advantage if a general roadmap existed for node introductions and the feature sizes those nodes would target. This would allow for the broad, simultaneous development of all the pieces of the puzzle required to bring a new node to market. For many years, the ITRS—the International Technology Roadmap for Semiconductors—published a general roadmap for the industry. These roadmaps stretched over 15 years and set general targets for the semiconductor market.

James Clarke, Intel:

> Intel was remarkably consistent [between roadmaps and results in quantum computing technologies], because we know how long it takes to develop a new technology. Even if we come out on a two-year cadence for transistor technology, the development time for those technologies is 10 to 12 years.

Boston Consulting Group and Semiconductor Industry Association, "Strengthening the Global Semiconductor Supply Chain in an Uncertain Era":

> The average length of time between when a new technological approach is introduced in a research paper and when it hits widescale commercial manufacturing is estimated to be about 10–15 years, but it could be much longer than that for scientific

Advance	Demonstration[/Focus]	Production
MOSFET	1959	1971 (first microprocessors)
Shallow trench isolation	∼Shibata et al. 1983	1997 (250 nm)
Strained silicon	1992/1998	2003
High-κ dielectric	1996/1998	∼ 2007 (45 nm)
Raised source/drain	1993/1998	2009
Multigate FET	1987/2000	2011
FinFET	∼Hisamoto et al. 1998/2000	2012–2013 (22 nm)
GAA	∼Lee et al. 2006	2024 (imec roadmap)
CFET	imec 2020	2032 (potential roadmap extension)

Table 1: First demonstrations and first use in large-scale production of major advances in front-end-of-line processes.

> breakthroughs that enable the current leading edge technologies. For example, Extreme Ultra-Violet (EUV) technology that is fundamental for the most advanced semiconductor manufacturing nodes took almost four decades from the early concept demos to its commercial implementation in fabs.

Table 1 shows a selection of major front-end-of-line advances, with dates for first demonstration and first production. Dates for strained silicon, high-κ dielectrics, raised source/drain, and multigate devices are from the 2022 IRDS report (Figure ES50), which separates early invention dates and the beginning of “focused research”.[22] The imec roadmap after GAA transistors goes out to 2036. For the remainder, I used my own somewhat arbitrary judgment for both the selection of advances and the interpretation of “first demonstration”, but this is the sort of thing people are thinking of when they say 10–15 years.

[22] I haven’t independently checked these dates, and specific references aren’t provided. I had previously done a cursory check for high-κ dielectric literature and had come up with 1999/2000. I think 1996/1998 also makes sense; there isn’t a single watershed result I can point to, and focused interest would have preceded publications by a year or two.

There are important related question about slack in that lead time. How many of those 10–15 years are for humans to take bathroom breaks? (Figuratively speaking.)

How much time do you need for serial experiments to solve problems of production at scale: process compatibility, thermal/mechanical/electrical integration, yield, throughput, aging? Is the time between demonstration and focused research required to select that technology from a field of candidates, or is the invention simply languishing until it’s relevant? Is there slack in industry coordination that a unitary AI would tighten? (Perhaps not as much as one might think—this was the point of extensive pre-competitive collaboration in the industry.)[23]

[23] See for example the quote from Joel Hruska above, or the successor to ITRS on its origin: “The goal of [the first International Roadmap for Semiconductors program in 1998] consisted of reducing the historical time of ∼25 years between major transistor innovations to less than half in order to save the semiconductor industry from reaching a major crisis.” IEEE 2022, Executive Summary, p. 66.

Speeding up this process is one of those problems where the more detail you know, the harder it automatically looks. My inside-view answers may be biased towards saying AI won't make much headway, since my head is in these details—but an AI or even future humans may find ways around them entirely, devoting great resources to doing so if necessary. Still, outside-view arguments that a pipeline bringing physical technology from concept to semiconductor-industry scale could be sped up by a factor of 10 are extremely unpersuasive to me.

Appendix C Nanomechanical computers

In Chapter 12 of *Nanosystems*, Eric Drexler "examines a representative set of components and subsystems for nanomechanical computers".[24] He describes a mechanical system of moving logic rods, sliding against a housing and pressing against one another, operating at room temperature, moving in a switching time of 0.1 ns, cycling at 1 GHz, dissipating around 1/600 of its mechanical energy per cycle, and suffering one error in every 10^{64} gates.

[24] Drexler 1992, p. 342.

I hope to convey how implausible this is from the perspective of modern nanomechanics. Intuition should suggest large error rates and much more dissipation. Below, I find tighter upper bounds on performance, with 10^{-3} error rates and dissipation per cycle comparable to the mechanical energy involved. These bounds use favorable calculations based on fundamental processes and relations ignored in *Nanosystems*; a more realistic estimate would suggest dissipation and noise worse by further orders of magnitude. Some calculations use the specifics of the exemplar logic system, but the same considerations apply generally to the molecular assemblies of *Nanosystems* in this operating regime.

Appendix C.1 Dissipation

The lowest-loss acoustic materials—defect-free bulk crystals and epitaxial films—might be able to keep dissipation that low at room temperature. In these systems, vibrations displace atoms in a highly regular crystal lattice of nearly ideal harmonic potentials (that is, the restoring force is nearly linear in displacement). A resonator can be engineered to confine these vibrations at a specific frequency—like a tuning fork or a quartz oscillator in a wristwatch—while allowing only a small fraction of the acoustic energy to be converted into other kinds of motion with each oscillation.

That dissipation is usually described in terms of a resonator's "quality factor", defined as $Q = 2\pi \times \frac{\text{energy stored}}{\text{energy lost per cycle}}$. A resonator that lost 1/600 of its energy per cycle would have a Q factor of $2\pi \times 600 \approx 3770$. Some loss mechanisms are specific to the resonator geometry, like acoustic radiation through points where the resonator is anchored to its substrate; some mechanisms originate in intrinsic material properties but can still be affected by resonator design, like thermoelastic damping; some are fully intrinsic, as the slight nonlinearity of interatomic forces allows acoustic excitations (phonons) to interact with one another locally. Non-resonant motion suffers attenuation by the same mechanisms and can be described by the Q equivalent to that attenuation.

These mechanisms and the limits they impose have been theoretically and experimentally studied in nanoscale mechanical systems, although not in the language of *Nanosystems*.[25] From this perspective, a simple fundamental limit on dissipation in the nanomechanical computer is at least two orders of magnitude worse, and limits in any practical device orders of magnitude further.

[25] For example, see Lifshitz and Roukes 2000; Cleland and Roukes 2002; Chandorkar et al. 2008; Ghaffari et al. 2013; Rodriguez et al. 2019; Bachtold, Moser, and Dykman 2022.

Limits to $f \cdot Q$ The damping mechanism first described by Akhiezer[26] is both intrinsic and local, originating from elastic nonlinearities, and thus generally insensitive to design. Akhiezer damping puts a fundamental limit on the product of Q and acoustic frequency f for a material at a given temperature:[27]

[26] Akhieser 1939.

[27] Cleland 2013, Chapter 8.

$$(f \cdot Q)_{\text{max}} \approx \frac{\rho v_\phi^2}{2\pi \gamma_{\text{eff}}^2 c_V T \tau_{\text{th}}}, \tag{1}$$

where ρ is material density, v_ϕ is the phase velocity of sound, c_V is the specific heat capacity at constant volume, τ_{th} is the phonon thermalization time, T is the temperature, and γ_{eff} is the effective Grüneisen parameter, which is a measure of the involved elastic nonlinearities intrinsic to the material. That nonlinearity means that the strain due to the acoustic mode at frequency f modulates the frequencies of thermal phonons already present, creating temperature variation between local phonon modes that quickly equilibrate by removing energy from the non-thermal strain field.[28] (This is in contrast to thermoelastic damping, which involves equilibration by thermal conduction between different spatial points.)

[28] If the oscillation is fast compared to the phonon thermalization time ($f\tau_{\text{th}} > 1$), you enter a different damping regime that can exceed this limit, but that would require something like operating at 1 THz—not relevant here.

For diamond at room temperature, that limit is about $Q \cdot f \leq 3.7 \times 10^{13}$ Hz.[29] If your resonator has a cycle time of 1 ns—a frequency of 1 GHz—then the maximum Q would be 37,000. Nothing about this process is specific to resonators. If you're moving acoustic energy at 10 GHz—as with a mechanical switch accelerating and stopping in 0.1 ns bursts—then the

[29] Chandorkar et al. 2008.

maximum equivalent Q at that frequency would be 3,700, and you'd lose $2\pi/Q \approx 1/600$ of your energy in 0.1 ns.

It's entirely coincidental how close this is to the calculated dissipation in the nanomechanical computer, because Drexler ignores Akhiezer damping in his calculations.[30]

[30] He does give a statement of a special case (using the "phonon viscosity" formulation first given by Mason 1960) in section 7.4.2, correctly observing that this case can be ignored for the vibrational modes of individual components. There's also a related dismissal in section 7.3.7, "Interfacial phonon-phonon scattering": "Nonlinear interactions permit phonons to scatter from one another, and the restoring forces in a nonbonded interface are substantially nonlinear.... A preliminary evaluation suggests that this effect is small, but a more thorough investigation would be desirable."

One's intuition ought to be that the nanomechanical logic system would at best dissipate on the order of all its stored energy every cycle, and probably would be overdamped.

The claimed performance would already be up against fundamental limits if the nanomechanical computer were a single defect-free isotopically pure diamond crystal. But, instead, it's made up of small components constrained by and colliding with one another under non-bonded interactions.

We should naturally consider the collective motion of those components—phonons in a material made of heavy "atoms" with van-der-Waals "bonds" between them. It looks like a molecular solid, and it's the motion of these molecules that constitutes the operation of the computer, not relatively long-lived phonons within the diamond components. As in an ordinary crystal lattice, it becomes necessary to consider the nearly continuous spectrum of collective vibrations of these components from very low frequencies to the highest frequencies relevant on operating timescales. The atoms are rather lumpy and the lattice somewhat amorphous, so abstracting the assembly into a solid yields a relatively optimistic treatment.

Regardless of how well the components fit together, the speed of sound is slower than that within the molecules,[31] the intermolecular forces are more nonlinear, and the heat capacity is larger (like that of a polymer, being made up of components each with many internal degrees of freedom). There's some room for specifics, but it's hard to make the accounting in Eqn. 1 come out with less than two orders of magnitude faster dissipation from phonon-phonon interactions alone.

[31] Consider a linear atomic chain with components like those in *Nanosystems* as (rigid) "atoms": masses of 10^{-21} kg, bond stiffnesses of 40 N/m, and (average) lattice spacing of 5 nm; the sound speed is $5\,\mathbf{nm} \times \sqrt{40\,(\mathrm{N/m})/10^{-21}\,\mathrm{kg}} = 1\,\mathrm{km/s}$, less than 1/10 that of bulk diamond.

That is, even if all other loss mechanisms could be engineered away without sacrificing all other design degrees of freedom, the energy lost to thermal structural vibrations through fundamental processes is worse than the purportedly conservative bounds in *Nanosystems*.

Molecular relaxation When the logic rod executes its motion, it is responding to changes in the equilibrium position $x_0(t) \sim -(1\,\mathrm{nm}) \cos \pi t/t_{\text{switch}}$ with respect to the alignment forces. If we move slowly, then the alignment forces do work equal to $\frac{1}{2} m v_{\max}^2$ when accelerating the rod, and opposite work decelerating the rod. The net work done is $\int_0^{t_{\text{switch}}} dt\, F \cdot v = \int_0^{\pi} d\phi\, mv_{\text{max}}^2 \sin\phi \times \cos\phi = 0$.

Because of the finite speed of sound in the rod, the center of mass sees an alignment force with a lag of about $\tau = \ell_{\text{rod}}/v_s = 6\,\text{ps} \approx 2t_{\text{switch}}/33$. That is, in the absence of dissipation and ignoring the smaller stretching stiffness of the rod, the equation of motion looks something like $m\ddot{x} = -k(x(t-\tau) - x_0(t))$. Because the force is slightly out of phase with the displacement, each switching motion does net work of roughly $\int_0^{t_{\text{switch}}} dt\, F \cdot v = \int_0^{\pi} d\phi\, mv_{\text{max}}^2 \sin\phi \times \cos(\phi + 2\pi/33) \approx 0.6 \times \frac{1}{2} mv_{\text{max}}^2$. If you didn't need the design to do anything else, this energy might be recoverable, but practically, this excess energy would be thermalized by friction or transmitted into structural vibrations (which are then thermalized).[32]

More importantly, work done by structural stress is a generic problem in a system with lag between stress and strain appreciable on operating timescales. That lag or relaxation time is the basic concept behind the standard (Zener) model of viscoelasticity and is equivalent to a stiffness with an imaginary (dissipative) part.[33] Again, the primary issue is not stress and strain within a component but rather the stress and strain in the molecular solid formed by the components. This alone would set a lower bound on dissipation around a single-digit fraction of the stored energy per cycle at the proposed operating rate. This mechanism is fundamental to any molecular assembly where the molecules do not behave like perfectly rigid bodies with respect to structural motion. A fix would involve operating so slowly or making components so small that the lag really is negligible.[34]

Edge friction and superlubricity *Nanosystems* does not consider friction due to edges of the contact area, which can dominate at the nanoscale. For example, one experiment found that inner atoms contributed frictional force within their experimental uncertainty of 2×10^{-2} fN/atom or 1 kPa, while edge atoms contributed 0.5–1 pN/atom depending on edge orientation relative to motion:[35]

> In conclusion, performing experiments in superlubric graphite-graphite contacts of different area and perimeter, we identified the contributions of the contact area and edges to the overall frictional force. We found that, per atom, the contribution of edge atoms is 4–5 orders of magnitude greater than that of inner atoms. The contact area is virtually frictionless, and for contact sizes below $\sim 10\,\mu$m, the total friction is determined mainly by

[32] The mass of the logic rod and the stiffness against its displacement imply fundamental frequencies around 70 GHz in isolation, so a linear chain of as few as 100 logic rods connected through sections of housing will have collective modes extending below 1 GHz that will be excited by the drive motion. (Higher frequency modes will also be excited because the drive force, being intermittent rather than purely sinusoidal, has spectral content at those frequencies, but less so.) Since the logic rods are by necessity stiffly mechanically coupled, there is no question of eliminating these modes by "[g]ood design practices" as Drexler suggests in the brief consideration he gives the problem in §12.3.4h.

[33] Cleland 2013, Chapter 8. Usually the Zener model is applied phenomenologically, but in this case the relaxation time has a very literal origin.

[34] I imagine making interfaces so stiff as to match the sound speeds of intercomponent and intracomponent motion within 1% would also work, but even if feasible it is almost certainly incompatible with other important design constraints.

[35] Qu et al. 2020.

the edges.

The experimental sliding speeds are much lower (by 7 orders of magnitude, in a limit where velocity dependence is logarithmic rather than linear) and the normal force much higher (4 OOM) than in *Nanosystems*. Still, for comparison, Drexler calculates a frictional force per unit area of 1 kPa in sleeve bearings with relative surface speeds of 1 m/s (again, roughly 0.02 fN/atom assuming carbon nanotubes). Of course, Drexler is considering diamondoid tubes, while the experiment uses graphene-graphene surfaces, arguably the paradigmatic superlubric system.

Usually, frictional force scales linearly with normal force and sliding speed; because this is (evidently) not always true at the nanoscale, it's difficult to evaluate a comparison like this. Still, edge atoms in the experiment contribute four orders of magnitude more friction per atom *in absolute terms* than the friction Drexler calculates, in a setting that would ordinarily be three net orders of magnitude more favorable even before considering material differences.

Optimistically, as few as 1% of interface atoms in *Nanosystems*' exemplar logic system are edge atoms. Even if sliding friction of contact area interiors were computed correctly, it would likely remain an irrelevant contribution to total friction.

Friction in superlubric systems has also been studied in simulation. While I've found it's difficult to tell from the outside how much a given simulation can really be trusted, the dominant contribution of edges is a consistent result. For example, Tangney, Louie, and Cohen 2004 perform molecular dynamics (MD) simulation of relative translation of concentric nanotubes, finding that "[t]he principal source of friction is found to be the ends of the tubes and hence dynamical friction is virtually independent of the overlap area between tubes"; Guo et al. 2011; Guo et al. 2012, and Koren and Duerig 2016 are also in agreement.

If we want to consider area contributions in isolation, Cook, Buehler, and Spakovszky 2013 critically reviews MD studies of area scaling of rotational friction in double-walled carbon nanotubes (DWNTs). The authors attempt to make sense of the finding that "the values for friction reported in the literature do not agree and show scatter over many orders of magnitude, which makes it difficult to use the data with confidence." They find reason to agree with most optimistic results, which yield about 0.5 fN/atom at a sliding speed of 100 m/s. That would imply 0.005 fN/atom at 1 m/s, assuming the linear scaling holds down to those speeds. The most pessimistic results are about 2.5 orders of magnitude higher; the only experimental result at the time estimated static friction at 2 orders of magnitude higher than the optimistic high-speed kinetic friction.

This scaled simulation result is in fact lower that *Nanosystems*' estimate for sleeve bearings by a factor of 4. Again, *Nanosystems* considers diamondoid bearings and interface spacing around 0.2 nm, while DWNTs are basically graphene-graphene interfaces with spacing around 0.35 nm (and buckle too easily to be used with the forces in *Nanosystems*). It's again difficult to interpret the comparison, except to say that the sleeve-bearing results are well into the superlubricity regime.[36] In any case, edge contributions should already be sufficient to overdamp logic rod motion.

[36] I'm not aware of any other diamond-diamond superlubricity results with any surface termination. It also seems *Nanosystems*' friction calculation gives much smaller numbers for DWNTs, making it hard to accept its formulas as giving an "upper bound" on drag (10.4.6f).

It may be possible to reduce friction beyond that of known superlubric systems through design or materials choice, but one would expect doing so to severely limit the design space.

Appendix C.2 Noise and error rates

The logic rod, as far as it loses energy to its environment, is conversely subject to random noise forces displacing it. This is a statement of the fluctuation-dissipation theorem. This is the fundamental relation that connects drag and Brownian motion in a fluid, or resistance and Johnson noise in an electrical circuit. *Nanosystems* does not seriously consider the effect of such noise forces on error rates, instead calculating the thermal excitation of the extensional degree of freedom of a logic rod plus some displacement into the limit stops at the ends of its range of motion, which is unsurprisingly irrelevant (a logical error rate of 10^{-64}).[37]

[37] Drexler 1992, §12.3.7, pp. 352–354.

A more appropriate calculation of error rates would consider the probability that thermal forces displace the rod by the threshold distance during a cycle. *Nanosystems* notes that such external bombardment occurs, although the extent of its consideration is a dismissal:[38]

[38] Drexler 1992, §5.3.1a, "The irrelevance of external bombardment".

> At equilibrium, however, an impinging gas molecule is as likely to absorb energy as to deliver it, and so molecular bombardment has no net effect on the amplitude of vibration. How a system is coupled to a thermal bath can affect its detailed dynamics (e.g., the smoothness or irregularity of its trajectory, the decay time for oscillations of unusual amplitude, etc.), but not the statistical distribution of dynamical quantities. This principle holds true for systems in general, and makes the study of po-

sitional uncertainty dependent only on potential energy functions.

This is another statement of the fluctuation-dissipation relation for the particular case where you only care about the variance in slow measurements of a single mode. It's correct (interpreted appropriately), but it is not license to treat a given degree of freedom as coupled only to the designed modes of vibration. We very much do care about detailed dynamics in mechanisms like the nanomechanical computer, with its switching time of 0.1 ns and a positional degree of freedom participating in a spectrum of modes at different frequencies.

For some intuition, note that the thermal average (root mean square) velocity of a mass m at temperature T is $v_{\text{thermal}} = \sqrt{k_B T/m}$. For the logic rod in the nanomechanical computer, this is 4.6 m/s. The RMS velocity of the exemplar rod during the switching motion is 11 m/s.[39] This raises the natural question: what prevents a mere 2σ excursion from causing an error?[40]

[39] Drexler 1992, Eqn. 12.10, p. 349.

[40] That is, why 10^{-64} rather than 10^{-2}?

The reply in *Nanosystems*, as far as it considers the question, is: "In a typical nanomechanical subsystem, a series of components is mechanically coupled, moving as a nearly rigid unit with respect to some motion coordinate q."[41] That is, the logic rod is tightly constrained to its trajectory by the surrounding mechanism; the net force that accelerates the rod $F_{\text{accel}} \approx 0.1$ nN is much smaller than the alignment force $F_{\text{al}} \approx 1$ nN that restores its position from excursions on the scale of 0.01 nm.

[41] Drexler 1992, §10.8, "Barriers in extended systems".

Specifically, in the exemplar system, a logic rod is confined in its housing with a stiffness against displacement of around 40 N/m. The logic rod itself is taken to have a stretching stiffness around 10 N/m between the gate and alignment knobs. The thermal average relative displacement between probe and gate at room temperature is then roughly 0.023 nm. Likewise, there may be thermal excursions against the 1 nN alignment force, changing the effective error threshold. This is more or less what goes into Drexler's calculation of error probability.[42]

[42] Drexler 1992, §12.3.7b.

In dynamical detail, an error occurs if the displacement of the logic rod is larger than the error threshold (0.7 nm) for the length of time it takes for a probe rod to pass (on the order of 10 ps). In the language of noise and measurement on finite timescales, we care about position noise in a bandwidth of around 100 GHz. In particular, the average (root mean square) displacement in that bandwidth should be much less than 0.7 nm. Quantitatively, if we target 0.023 nm, then in terms of noise power spectral density, the requirement is $S_{xx}(\omega) = (0.023\,\mathrm{nm})^2/(2 \cdot 100\,\mathrm{GHz})$.[43]

The complementary picture looks at fluctuations in forces. If environmental bombardment applies a nN-scale force for this brief time, then the rod can be displaced against the alignment force, causing an error. How often does that happen?

[43] Using white noise to be conservative, and using the two-sided spectral density with the 2π convention such that for white noise with variance σ^2 the spectrum is $S(\omega) = \sigma^2$ (following Clerk et al. 2010). This discussion will be very informal, but the point only needs orders of magnitude to carry.

Quantum limits *Nanosystems* notes a number of times that the systems it describes are far from quantum limits.[44] Above, the purported dissipation was found to be at the (room temperature) quantum limit for bulk diamond and well below that for a molecular solid. For some more intuition about how plausible *Nanosystems*' numbers are, let's consider limits on noise.

[44] E.g., Chapter 5, "Positional Uncertainty".

According to the fluctuation-dissipation relation, the (two-sided) power spectral density of the noise force at frequencies $\omega \ll k_B T/\hbar$, given a damping rate γ, is the white spectrum $S_{FF}(\omega) = 2m\gamma k_B T$. For the sliding friction of the logic rod calculated in *Nanosystems* Section 12.3.4c, the damping rate can be calculated from the average velocity and implied frictional force as $\gamma = 2\pi \times 4\,\mathrm{MHz}$, so $S_{FF}(\omega) \approx 4 \times 10^{-5}\,\mathrm{fN}^2/\mathrm{Hz}$.

We can compare this to the quantum-limited force noise due to the uncertainty principle, which in this context tells us that $S_{xx} S_{FF} \geq \hbar^2/4$.[45] Together with the position uncertainty above, the minimum uncertainty in the alignment force is $S_{FF}(\omega) \geq 4 \times 10^{-5}\,\mathrm{fN}^2/\mathrm{Hz}$. Again, there's no physical connection between the two calculations. It's just coincidental that the given friction adds force noise equivalent to the minimum force noise arising from quantum uncertainty. Experiments have, at least, measured noise this low, although at lower frequencies and in limited bandwidths, for example in LIGO and other experiments going to great lengths (reaching ultra-high vacuum and extreme cryogenic temperatures, engineering structures to do nothing but block phonon transmission at certain frequencies, harnessing quantum correlations) to reach those sensitivities, sometimes just to show that they can (e.g., Mason et al. 2019).[46]

[45] Braginsky, Khalili, and Thorne 1992.

[46] I'm emphasizing this in an attempt to transmit some perspective on how implausible that noise level is for a room-temperature device with moving parts at gigahertz frequencies where everything is in contact with everything else.

It at least remains true that these systems should be orders of magnitude away from quantum limits. The noise in the restoring force associated with the 0.023 nm thermal displacements is about $40\,\mathrm{N/m} \times 0.023\,\mathrm{nm} \approx 1\,\mathrm{nN}$. The timescale for restoring the position is on the order of $10\,\mathrm{ps}$. The rod applies an equal and opposite force on the alignment stop. Every comparable component in the assembly is experiencing similar forces, which are transmitted throughout the structure because everything is stiffly mechanically coupled. Informally, it shouldn't surprise us to find noise forces much larger than 1 nN on 10 ps timescales in a realistic model.

Force noise estimate If the numbers in *Nanosystems* are not realistic, then just how large are fluctuations in force on sub-nanosecond timescales? We can at least find a lower bound based on the minimum dissipation we've identified.

Using the friction in *Nanosystems* and Akhiezer damping in bulk diamond, the noise from these two broadband sources alone would be

$$S_{FF}(\omega) = 2mk_BT\left(2\pi \times 4\,\mathrm{MHz} + \frac{\omega^2}{2\pi \times 3.7 \times 10^{13}\,\mathrm{Hz}}\right) \quad (2)$$

$$\sigma_F^2 = \frac{1}{2\pi}\int_{-2\pi\times 100\,\mathrm{GHz}}^{2\pi\times 100\,\mathrm{GHz}} d\omega S_{FF}(\omega) \quad (3)$$

$$\sigma_F \approx 14\,\mathrm{pN}. \quad (4)$$

where the variance σ_F^2 is taken over 10 ps averages of the force. This would plausibly be fine if correct and complete, because it's small compared to the alignment force.

If actual damping is a conservative three orders of magnitude larger, then the average noise force is nearly $0.5\,\mathrm{nN}$, overcoming the constant-force alignment spring by enough to cause an error in 0.1–1% of cycles.[47] This is already enough to make nanometer-precision, GHz-frequency operation of this sort of assembly impractical at room temperature.

[47] This is easily consistent with the originally calculated positional uncertainty of 0.023 nm, since 0.5 nN/(40 N/m) is still only half that. That's not to say much larger noise forces aren't possible (even likely), but rather just that they're approaching an overdamped Brownian motion regime that looks somewhat different.

It's important to note that most problems I've identified can be addressed by making the device both slower and larger. (Going slower has to include reducing alignment forces, or else errors can happen in the same short time and the noise bandwidth that you're sensitive to doesn't go down. But the noise force amplitude related to viscous damping scales with the square root of that bandwidth and so becomes larger relative to the alignment forces. So you have to make things larger, too.) Going to cryogenic temperatures also helps with some things, but increases friction, reduces thermal conductivity, and increases sensitivity to local heating. I suspect a workable design would not be competitive with field-effect transistors even in principle, but that's a hard argument to make airtight. A fix involving scaling would also be less helpful for the various other assemblies that require sub-nanometer precision in itself for chemistry.

The sketch above does not account for the contribution of any resonant modes involving relative displacement of the logic knobs, which cannot be eliminated in a stiffly-coupled assembly of many parts (consider note 32

above). It also ignores the response time of alignment forces, which is comparable to the time required for an erroneous knob passage. Noise added by mechanical amplification (in particular anywhere signals fan out) is another critical subject with fundamental consequences left fully open.

Appendix C.3 Errata?

Finally, it's worth looking more closely at some calculations in *Nanosystems*. Drexler uses the physics of dislocations in a crystal lattice as a model for many calculations of friction between nanoscale parts. The idea is resourceful given the tools available at the time, but it's also not clear how appropriate that model is, and some of his terminology (e.g., "band-stiffness scattering", "band-flutter scattering", "shear-reflection drag") does not appear in the literature before or after *Nanosystems*. Even once nanotribology coalesced in the early 1990s as a field systematically studying friction and interfacial mechanics at the atomic scale, it did not draw on the methods of *Nanosystems* or its references.[48] One would not have expected these methods to be complete or correct, and we have found that they are not.

[48] Bhushan, Israelachvili, and Landman 1995.

That said, I believe there are also problems in their application on their own terms. For example, *Nanosystems* Section 7.3.5, "Scattering from alignment bands in bearings", describes calculations for drag between sliding surfaces with some analogy to dislocation physics. A key result quoted from a personal communication is a dimensionless phonon transmission coefficient, Eqn. 7.40:

$$T_{\text{trans}} = \frac{4}{3}\frac{\int_0^1 \frac{k^3}{e^{k/T'}-1}\int_0^{\pi/2}\frac{\sin 2\theta}{(d'k\cos\theta)^2+4}\,d\theta\,dk}{\int_0^1 \frac{k^3}{e^{k/T'}-1}dk}, \tag{5}$$

where $T' = T/T_D$ and $d' = (6\pi^2 n_V)^{1/3}M/k_a$ are dimensionless parameters, T is the temperature, T_D is the Debye temperature of the medium, n_V is the number density of atoms in the medium, M is the elastic modulus of the medium, and k_a is the stiffness per unit area of the interface. Friction, as calculated in *Nanosystems*, scales with T_{trans}.

This is followed by the approximation (Eqns. 7.41 and 7.42), corresponding to the dotted lines in Figure 1:

$$T_{\text{trans}} \approx \frac{z}{1+3z}; \qquad z = 0.6d_{\text{n}}^{-1.7}(1+0.075T'^{-1.8}) \tag{6}$$

$$T_{\text{trans}} \approx z, \qquad z \ll 1 \tag{7}$$

where $d_{\text{n}} = n_V^{1/3}M/k_a = d' \times (6\pi^2)^{-1/3}$.[49]

[49] I'm assuming that the $-1/3$ in the book's definition $d_{\text{n}} = n_V^{-1/3}M/k_a$ is a typographical error, since d_{n} needs to be dimensionless. Similarly the Debye wavenumber is $(6\pi^2 n_V)^{1/3}$ rather than the $(6\pi n_V)^{1/3}$ of the book's Eqn. 7.29, since that's the standard definition and it produces the matching figure.

By the logic sketched in the text, the $\sin 2\theta$ in the numerator of Equation 5 should instead be $\sin\theta$ (from integrating in spherical coordinates). The equation as written implies an extra factor of $2\cos\theta$ with no justification and contradicts the comment on the next page that grazing-incidence ($\theta = \pi/2$) phonons make a large contribution to T_{trans}. Unfortunately, the reference is to personal communications with J. Soreff (1991) and there is no relevant published work by the same author. Changing $\sin 2\theta$ to $\sin\theta$ increases T_{trans} (and consequent drag power) by a factor of 10 to 100 for the sets of parameters used in the text. In Figure 1 we reproduce Fig. 7.6 in *Nanosystems* with the formula as written, alongside a comparison with the $\sin\theta$ version.

Appendix C.4 Meditation

I didn't keep track, but I likely spent on the order of 100 hours reading and digesting *Nanosystems* and related texts,[50] identifying problems in the analysis, estimating corrections, and choosing parts of all that to work into a critique of in-principle feasibility spelled out in a way that might be persuasive to non-experts. This EA Forum comment outlines my thoughts early in this process, including intuitive bullet versions of some of my arguments above, among others. Somewhat later, I expanded on some particulars, alluding to back-of-the-envelope calculations along the same lines. This final product is narrow in scope and considers a handful of dissipation and noise mechanisms applied to about 11 pages in Chapter 12 of *Nanosystems*, but the chosen claims—considerations that apply to the systems described throughout the book—are argued quantitatively in some detail.

[50] These include Freitas 1999; Freitas 2003; Phoenix, Moriarty, and Jones 2004–2005; Jones 2004–; leplen 2013; Beckstead 2015; Marblestone 2021; Snodin 2022; bhauth 2023, and GiveWell's non-verbatim summaries of conversations with Adam Marblestone, Chris Phoenix, Eric Drexler, and Philip Moriarty.

If I'd only cared about persuading myself, I would have stopped much sooner. I did, in fact. I spent a weekend or two with *Nanosystems* as a student and set it down with very high confidence that its nanomechanical proposals were unworkable and high confidence that nothing in that ballpark was workable in principle. The discussion above just sharpens a couple of the basic intuitions and physical arguments about nanomechanics I had relied on as a student without having thoroughly checked them—and without having identified particular errors in *Nanosystems*. At that point, from my perspective as a researcher, there was no benefit to further engagement on the subject. It would not result in anything more publishable than a blog post or yield anything like a research program.

I suspect nearly all scientists who engage with *Nanosystems* stop there, since the level of public discussion is generally around the level of my early

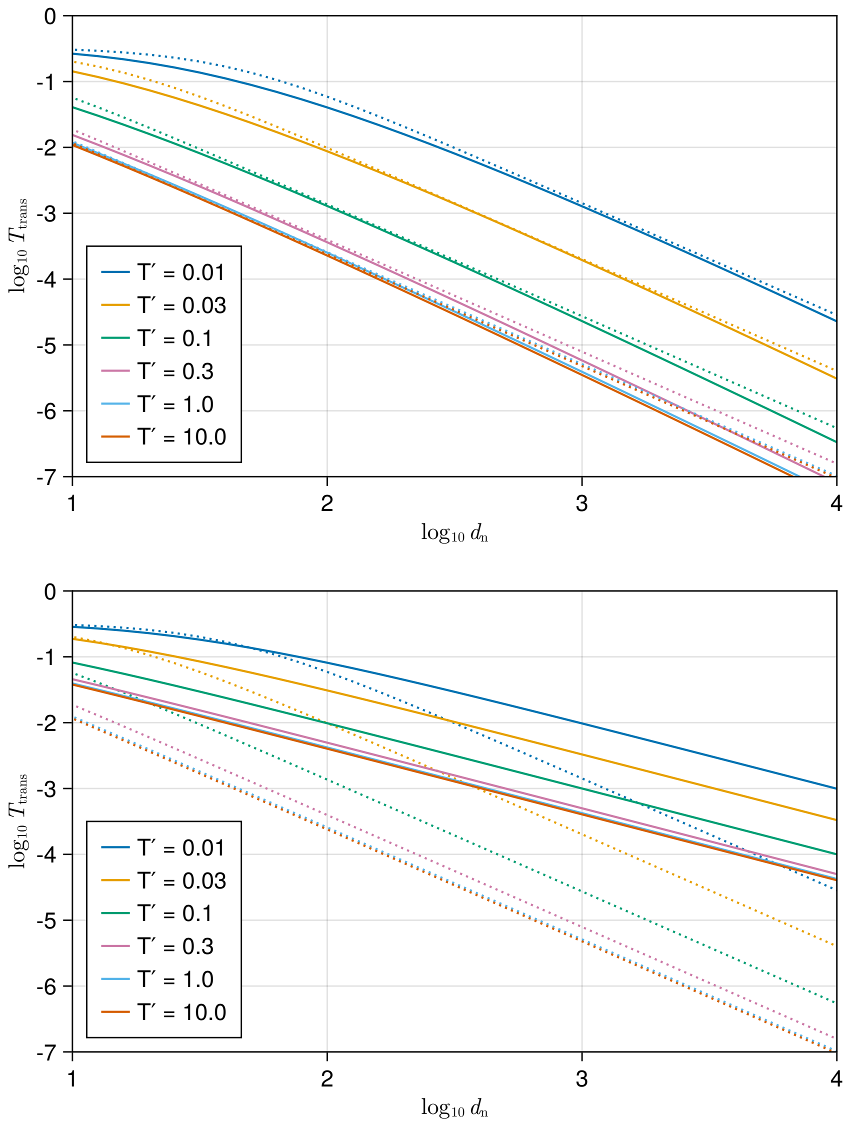

Figure 1: Phonon transmission coefficient at a compliant interface T_{trans}. Friction, as calculated in *Nanosystems*, scales linearly with T_{trans}. Top: Identical to Fig. 7.6 in *Nanosystems*, produced using our own implementation of its equations. Solid lines use Equation 5, dotted lines use the approximation Equation 6. Bottom: The same plot, but (from my understanding) corrected, meaning that the solid lines use $\sin\theta$ in the spherical integral rather than $\sin 2\theta$. The Debye temperature for diamond used by *Nanosystems* gives $T' = 0.13$, and key exemplars in *Nanosystems* have $\log_{10} d_{\text{n}}$ ranging roughly between 2 and 3.

intuitions and arguments—closer to blog posts than to problem sets. Even that level of public engagement is seen as somewhat embarrassing, with Smalley as the obvious case.

So detailed critiques are absent from the published literature, but predictably so. It's time-consuming to review even a small facet of a text that jumps so breezily from one problem to another, and there's little to gain. There's even risk of backfiring—my anecdotal impression is that lower-effort critiques have historically had mixed reception among non-experts, resulting in claims that *Nanosystems* provides a detailed blueprint that has survived review, or that the best that critics can do is engage with a straw man, describe mere engineering constraints, or regulate social status.

Having now followed through somewhat on my early intuitions, I also have more confidence in the objections others have raised on the subjects further from my expertise. I expect many intuitive critiques would carry with generality if elaborated, while others identify practical difficulties that considered together produce unsatisfiable engineering constraints.[51]

[51] For example, this goes for the class of "common intuitive objections" I list at the beginning here and for those bhauth gives here. Not every one applies to every possible nanosystem, but that's not necessary to form confident beliefs based on diverse particulars with wide total coverage.

I worry that I've done some injustice in taking everything but simple nanomechanics as given, suggesting that anything I don't mention or use without objection is correct, or by treating the proposed systems with seriousness out of proportion to the work put in by *Nanosystems* itself. Similarly, I don't want to give the impression that I speak for the scientific community, or that this is the best this angle of criticism has to offer. I've undoubtedly missed things—some favorable to my arguments, some not. I don't have any special claim to authority, and there's nothing novel here. I'm just a person whose scientific background is tangential to some topics in *Nanosystems*.

I've tried to write this appendix in a measured way, avoiding going beyond claims that I directly support in the text. Still, I should state clearly that the calculations above are a modest contributor to my own high confidence that claims in *Nanosystems* and its subsequent tradition should not particularly factor into how we think about existential risk or the future of computing. From my perspective, the argument that a particular system is unworkable is foremost evidence that the level of analysis in *Nanosystems* is not nearly sufficient to inform "lower bounds" on possible or plausible technological capabilities. This appendix has described, even in a quantitative sense, the least of the book's problems. They are just the simplest and most general considerations I've identified that can (I hope) be communicated forcefully while relying minimally on implicit expert knowledge.

Appendix D Predictions

While I forecast as a hobby, I'm generally disinclined to give probabilities when no decision hinges on them. In this case, because the essay body attempts to avoid weighing evidence while suggesting places to look for evidence, it seems important to provide some context on my own views for the sake of transparency.

1. In 10 years, I judge that the weight of evidence is against similarity to Scenario 1: 50%
2. In 10 years, I judge that the weight of evidence is against similarity to Scenario 2: 5%
3. An open-source model competitive with GPT-4 is available by the end of 2024: 60%
4. GPT-4 is obsolete by 2028: 80%
5. OpenAI revenues exceed expenses by 2028: 40%
6. Leading-edge AI hardware operating principles are discontinuous with field-effect transistors (e.g., optical computing, spins, magnons, plasmons, flux quanta, nanomechanics)...
   (a) ... by 2043: 1%
   (b) ... by 2070: 5%
7. Leading-edge AI hardware incorporates some "beyond-CMOS" technologies like the above integrated with traditional CMOS designs...
   (a) ... by 2043: 30%
   (b) ... by 2070: 80%
8. A major change in transistor structure, comparable to the change from planar to fin FETs, goes from initial invention to leading-edge process nodes in under 1 year...
   (a) ... by 2043: $\ll$1%
   (b) ... by 2070: $<$1%
9. A major change in transistor structure, comparable to the change from planar to fin FETs, goes from initial invention to leading-edge process nodes in under 5 years...

   (a) ... by 2043: 1%
   (b) ... by 2070: 5%

10. The next major change in AI architecture, comparable to the change from RNNs to transformers, goes from initial invention to leading-edge models in under...
    (a) 1 year: 5%
    (b) 5 years: 50%
11. Nanomechanical computers with the operating characteristics described in *Nanosystems* are feasible in principle: $<$1%
12. Nanomechanical computers are competitive with leading-edge transistor-based computers by 2070: $\ll$1%
13. A peer with similar background agrees that I accurately describe generic obstacles to competitive nanomechanical computation as described in *Nanosystems*: 95%
14. Eric Drexler agrees that I accurately describe generic obstacles to competitive nanomechanical computation as described in *Nanosystems*: 10%
15. Eric Drexler agrees that friction calculations in *Nanosystems* are not accurate upper bounds: 40%

References

Aghion, P., B. F. Jones, and C. I. Jones (2017). "Artificial Intelligence and Economic Growth". (Cit. on p. 14).

Akhieser, A (1939). "On the absorption of sound in solids". *J. Phys.(Ussr)* **1** 277. (Cit. on p. 18).

Bachtold, A., J. Moser, and M. I. Dykman (2022). "Mesoscopic physics of nanomechanical systems". *Reviews of Modern Physics* **94** 045005. DOI: `10.1103/RevModPhys.94.045005`. (Cit. on p. 18).

Beckstead, N. (2015). "Risks from Atomically Precise Manufacturing". URL: `https://www.openphilanthropy.org/research/atomically-precise-manufacturing/`. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

bhauth (2023). "grey goo is unlikely". URL: https://www.lesswrong.com/posts/FijbeqdovkgAusGgz/grey-goo-is-unlikely. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Bhushan, B., J. N. Israelachvili, and U. Landman (1995). "Nanotribology: friction, wear and lubrication at the atomic scale". *Nature* **374** 607–616. DOI: 10.1038/374607a0. (Cit. on p. 26).

Braginsky, V. B., F. Y. Khalili, and K. S. Thorne (1992). *Quantum Measurement*. Cambridge University Press. DOI: 10.1017/CBO9780511622748 (cit. on p. 24).

Carlsmith, J. (2020). "How Much Computational Power Does It Take to Match the Human Brain?" URL: https://www.openphilanthropy.org/research/how-much-computational-power-does-it-take-to-match-the-human-brain/. [Accessed 2023-05-30 (archive)] (cit. on p. 10).

— (2021). *Is Power-Seeking AI an Existential Risk?* en. arXiv:2206.13353 [cs] (cit. on p. 10).

Chandorkar, S. A., M. Agarwal, R. Melamud, R. N. Candler, K. E. Goodson, and T. W. Kenny (2008). "Limits of quality factor in bulk-mode micromechanical resonators". *2008 IEEE 21st International Conference on Micro Electro Mechanical Systems*. ISSN: 1084-6999, pp. 74–77. DOI: 10.1109/MEMSYS.2008.4443596 (cit. on p. 18).

Cleland, A. N. and M. L. Roukes (2002). "Noise processes in nanomechanical resonators". *Journal of Applied Physics* **92** 2758–2769. DOI: 10.1063/1.1499745. (Cit. on p. 18).

Cleland, A. (2013). *Foundations of Nanomechanics: From Solid-State Theory to Device Applications*. Advanced Texts in Physics. Springer Berlin Heidelberg. ISBN: 9783662052877 (cit. on pp. 18, 20).

Clerk, A. A., M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf (2010). "Introduction to quantum noise, measurement, and amplification". *Reviews of Modern Physics* **82** 1155–1208. DOI: 10.1103/RevModPhys.82.1155. (Cit. on p. 24).

Cook, E. H., M. J. Buehler, and Z. S. Spakovszky (2013). "Mechanism of friction in rotating carbon nanotube bearings". *Journal of the Mechanics and Physics of Solids* **61** 652–673. DOI: 10.1016/j.jmps.2012.08.004. (Cit. on p. 21).

Cotra, A. (2020). "Forecasting TAI with biological anchors (draft report)". URL: https://drive.google.com/drive/u/0/folders/

15ArhEPZSTYU8f012bs6ehPS6-xmhtBPP. [Accessed 2023-05-30] (cit. on pp. 1, 10).

Davidson, T. (2021a). "Could Advanced AI Drive Explosive Economic Growth?" URL: https://www.openphilanthropy.org/research/could-advanced-ai-drive-explosive-economic-growth/. [Accessed 2023-05-30 (archive)] (cit. on pp. 1–3, 10).

— (2021b). "Semi-informative priors over AI timelines". URL: https://www.openphilanthropy.org/research/semi-informative-priors-over-ai-timelines/. [Accessed 2023-05-30 (archive)] (cit. on p. 10).

— (2023). "What a compute-centric framework says about AI takeoff speeds (draft report)". URL: https://docs.google.com/document/d/1os_4YOw6Xv33KjX-kR76D3kW1drkWRHKG2caeiEWzNs. [Accessed 2023-05-30] (cit. on pp. 1, 2, 10, 12).

Drexler, K. E. (1992). *Nanosystems: Molecular Machinery, Manufacturing, and Computation*. John Wiley & Sons (cit. on pp. 17, 22, 23).

Freitas Jr., R. A. (1999). *Nanomedicine, Volume I: Basic Capabilities*. Landes Bioscience (cit. on p. 27).

— (2003). *Nanomedicine, Volume II: Biocompatibility*. Landes Bioscience (cit. on p. 27).

Ghaffari, S., S. A. Chandorkar, S. Wang, E. J. Ng, C. H. Ahn, V. Hong, Y. Yang, and T. W. Kenny (2013). "Quantum Limit of Quality Factor in Silicon Micro and Nano Mechanical Resonators". *Scientific Reports* **3** 3244. DOI: 10.1038/srep03244. (Cit. on p. 18).

Grace, K. (2022). "Counterarguments to the basic AI x-risk case". URL: https://aiimpacts.org/counterarguments-to-the-basic-ai-x-risk-case/. [Accessed 2023-05-30 (archive)] (cit. on p. 14).

Guo, Z., T. Chang, X. Guo, and H. Gao (2011). "Thermal-Induced Edge Barriers and Forces in Interlayer Interaction of Concentric Carbon Nanotubes". *Physical Review Letters* **107** 105502. DOI: 10.1103/PhysRevLett.107.105502. (Cit. on p. 21).

— (2012). "Mechanics of thermophoretic and thermally induced edge forces in carbon nanotube nanodevices". *Journal of the Mechanics and Physics of Solids* **60** 1676–1687. DOI: 10.1016/j.jmps.2012.04.013. (Cit. on p. 21).

Hisamoto, D., Wen-Chin Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, Tsu-Jae King, J. Bokor, and Chenming Hu (1998). "A folded-channel MOSFET for deep-sub-tenth micron era". en. *International Electron De-*

vices Meeting 1998. Technical Digest (Cat. No.98CH36217). San Francisco, CA, USA: IEEE, pp. 1032–1034. ISBN: 978-0-7803-4774-8. DOI: 10.1109/IEDM.1998.746531 (cit. on p. 16).

IEEE (2022). "International Roadmap for Devices and Systems: 2022 Edition". URL: https://irds.ieee.org/editions/2022 (cit. on p. 16).

Jones, B. (2021). "Comments on "Could Advanced AI Drive Explosive Economic Growth?"" URL: https://docs.google.com/document/d/1jP9Bb6J6BXH5v6EshsPF2NE1GiWatPxUUrK9wDEpTqA/ (cit. on pp. 3, 12).

Jones, R. (2004–). "Soft Machines, Category: Radical nanotechnology and MNT". URL: http://www.softmachines.org/wordpress/?cat=3. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Koren, E. and U. Duerig (2016). "Moiré scaling of the sliding force in twisted bilayer graphene". *Physical Review B* **94** 045401. DOI: 10.1103/PhysRevB.94.045401. (Cit. on p. 21).

Lee, H., L.-E. Yu, S.-W. Ryu, J.-W. Han, K. Jeon, D.-Y. Jang, K.-H. Kim, J. Lee, J.-H. Kim, S. Jeon, G. Lee, J. Oh, Y. Park, W. Bae, H. Lee, J. Yang, J. Yoo, S. Kim, and Y.-K. Choi (2006). "Sub-5nm All-Around Gate FinFET for Ultimate Scaling". *2006 Symposium on VLSI Technology, 2006. Digest of Technical Papers.* ISSN: 2158-9682, pp. 58–59. DOI: 10.1109/VLSIT.2006.1705215 (cit. on p. 16).

Leiserson, C. E., N. C. Thompson, J. S. Emer, B. C. Kuszmaul, B. W. Lampson, D. Sanchez, and T. B. Schardl (2020). "There's plenty of room at the Top: What will drive computer performance after Moore's law?" *Science* **368** eaam9744. DOI: 10.1126/science.aam9744. (Cit. on p. 8).

leplen (2013). "How probable is Molecular Nanotech?" URL: https://www.lesswrong.com/posts/ZeDS9kCSzGwsxCKBW/how-probable-is-molecular-nanotech. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Lifshitz, R. and M. L. Roukes (2000). "Thermoelastic damping in micro- and nanomechanical systems". *Physical Review B* **61** 5600–5609. DOI: 10.1103/PhysRevB.61.5600. (Cit. on p. 18).

Marblestone, A. (2021). "Bottleneck analysis: positional chemistry". URL: https://web.mit.edu/amarbles/www/docs/Bottleneck_analysis_positional_chemistry.pdf. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Markov, I. L. (2014). "Limits on Fundamental Limits to Computation". *Nature* **512** 147–154. DOI: 10.1038/nature13570. (Cit. on p. 14).

Mason, D., J. Chen, M. Rossi, Y. Tsaturyan, and A. Schliesser (2019). "Continuous force and displacement measurement below the standard quantum limit". *Nature Physics* **15** 745–749. DOI: 10.1038/s41567-019-0533-5. (Cit. on p. 24).

Mason, W. P. (1960). "Phonon Viscosity and Its Effect on Acoustic Wave Attenuation and Dislocation Motion". *The Journal of the Acoustical Society of America* **32** 458–472. DOI: 10.1121/1.1908099. (Cit. on p. 19).

Mody, C. (2016). *The Long Arm of Moore's Law: Microelectronics and American Science*. MIT Press (cit. on p. 3).

Phoenix, C., P. Moriarty, and R. Jones (2004–2005). "Phoenix-Moriarty debate". URL: http://www.softmachines.org/wordpress/?p=70. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Qu, C., K. Wang, J. Wang, Y. Gongyang, R. W. Carpick, M. Urbakh, and Q. Zheng (2020). "Origin of Friction in Superlubric Graphite Contacts". *Physical Review Letters* **125** 126102. DOI: 10.1103/PhysRevLett.125.126102. (Cit. on p. 20).

Rodriguez, J., S. A. Chandorkar, C. A. Watson, G. M. Glaze, C. H. Ahn, E. J. Ng, Y. Yang, and T. W. Kenny (2019). "Direct Detection of Akhiezer Damping in a Silicon MEMS Resonator". *Scientific Reports* **9** 2244. DOI: 10.1038/s41598-019-38847-6. (Cit. on p. 18).

Roser, M. (2020). "Why did renewables become so cheap so fast?" URL: https://ourworldindata.org/cheap-renewables-growth. [Accessed 2023-05-30 (archive)] (cit. on p. 6).

Shibata, T., R. Nakayama, K. Kurosawa, S. Onga, M. Konaka, and H. Iizuka (1983). "A simplified box (buried-oxide) isolation technology for megabit dynamic memories". *1983 International Electron Devices Meeting*, pp. 27–30. DOI: 10.1109/IEDM.1983.190432 (cit. on p. 16).

Snodin, B. (2022). "My thoughts on nanotechnology strategy research as an EA cause area". URL: https://forum.effectivealtruism.org/posts/oqBJk2Ae3RBegtFfn/my-thoughts-on-nanotechnology-strategy-research-as-an-ea. [Accessed 2023-05-30 (archive)] (cit. on p. 27).

Sotala, K. (2018). "Disjunctive scenarios of catastrophic AI risk". *Artificial intelligence safety and security*. Chapman and Hall/CRC, pp. 315–337 (cit. on p. 14).

Tangney, P., S. G. Louie, and M. L. Cohen (2004). “Dynamic Sliding Friction between Concentric Carbon Nanotubes”. *Physical Review Letters* **93** 065503. DOI: 10.1103/PhysRevLett.93.065503. (Cit. on p. 21).