The Bounded Parallelizability of R&D: Theory and Application to AI

Philip Trammell[1]

June 9, 2026

Technological development is typically modeled as Cobb-Douglas in (a) the quality of existing technology and (b) the quantity of research inputs, such as researchers or lab equipment. In fact, technology and research inputs are often gross complements: even infinite quantities of old equipment cannot drive progress as quickly as finite quantities of new tools which allow more of the research process to occur simultaneously. I document this and consider its implications for the automation of research in AI. If relieving parallelization constraints is no harder than growing the stock of effective research inputs, then automating AI R&D will cause AI development to accelerate almost as quickly as in the Cobb-Douglas case. Otherwise, parallelization constraints will eventually bind, and automating AI R&D may have little impact on the speed of AI development in the long run.

[1]Epoch AI and Digital Economy Lab, Stanford University. Contact: trammell@stanford.edu.

1 Introduction

Fixing our technology, suppose we double our capital, labor, land, and any other rival inputs to production. The standard "replication argument" suggests that output should at least double: at worst, we can always do the same thing twice.

By contrast, if we double the capital, labor, and so on used in research and development, we have no such reason to expect that R&D will proceed twice as quickly. This is because the output of R&D, namely new technology, is nonrival. The replication argument tells us that at worst we can double the number of times we write the same papers, say, or produce the same blueprints for new products. But having two copies of the same paper or blueprint is no more valuable than having one. Increasing the inputs to R&D only accelerates technological development to the extent that R&D can be *parallelized*, so that we can use the abundant inputs to work on different useful research tasks simultaneously.

The current approach: unbounded parallelizability. Let A index the state of some technology. It may represent total factor productivity (TFP) throughout an economy, for example, or TFP within some industry, the number of floating-point operations per second we can produce from a gram of silicon, or the number of times we can perform some task using a given number of floating-point operations.

In Romer (1990), Aghion and Howitt (1992), Jones (1995), and almost every other growth model, A's growth rate at t is assumed to exhibit a constant elasticity in the research inputs R_t, such as researchers and lab equipment, dedicated to increasing A at t. In particular, it is assumed that

$$g_{At} \propto A_t^{-\beta} R_t^{\lambda} \tag{1}$$

for some value of β (positive, if "ideas get harder to find"; 0 if not) and λ.

The λ term, typically assumed to be in $(0, 1]$, governs the extent to which larger quantities of research inputs can be used to make technological progress simultaneously. If λ equals 1, doubling research inputs doubles the rate of progress. If less, less. In any case, the elasticity of $\dot{A}$ to R is assumed to be positive and constant, so R&D can be parallelized unboundedly: with enough inputs, technology advances arbitrarily quickly.

This paper: technology-bounded parallelizability. If the parallelizability of R&D is unbounded, arbitrarily fast progress can be made on any technological frontier with a large enough quantity of the research inputs available *at that frontier*. This is implausible.

Suppose A denotes the productivity of an assembly line producing a particular model of robot. The relevant research inputs we have today are human engineers, managers, and the equipment (including today's robotic equipment) that they use to design and implement more efficient robot production processes. If we scaled these inputs indefinitely, the rate of growth in the assembly line's productivity would rise only a finite amount. Our research inputs can multiply until the factory floor is packed with engineers like a jar of olives, and a trillion more on Zoom have joined the throng, and still the efficiency of the assembly line will not double in ten minutes.

The bottleneck explored here is a lack of "parallelization technology", not a hard cap on technological progress per unit time. Sufficiently advanced technology may allow R&D to proceed much more rapidly than is possible with current technology. If we already knew how to build fast-moving, superintelligent, and highly coordinated robotic engineers, a large enough number of them might well take less than ten minutes to dismantle our original assembly line and devise and build a new one producing the same old robot model twice as quickly. It is also true that technological development may require time that no technology can save, e.g. because of the need for serial experiments or computations, and physical bounds on the speed at which these can be performed; but this paper is about the case, much more relevant to most applications today, in which better technology of some kind could let us do R&D more quickly.

In sum, I will propose a model in which technology itself is a gross complement to other research inputs in the R&D process:

$$g_{At} \propto A_t^{-\beta} f(A_t, R_t), \quad \beta > 0, \tag{2}$$

where $\lim_{R\to\infty} f(A, R)$ is finite but increasing in A. I assume $\beta > 0$ both for realism (Bloom et al., 2020) and because this is the case in which sustained exponential growth requires ever-growing research inputs, so constraints on parallelizability may bind.

Outline. Section 2 introduces a simple, one-sector version of the model. Section 3 introduces a multi-sector version with technology spillovers between sectors. Section 4 documents how the model captures important constraints on the AI R&D process neglected by standard models in which research can be parallelized unboundedly. Section 5 concludes by discussing the implications of this finding for the future of AI and perhaps of growth more broadly.

Related literature. On the growth theory front, there are four existing models in which the speed of technological development along some dimension can increase less than power-functionally in research inputs. The four are very similar to each other: in each case, the technology's growth takes the form

$$g_{At} \propto f(R_t) \tag{3}$$

for some $f(\cdot)$: that is, the technology's growth is "endogenous" ($\beta = 0$, so that constant research inputs sustain exponential growth), and the parallelizability of R&D does not depend on technology at all. In particular,[1]

- in Lashkari (2023) and Aghion et al. (2025), $f(\cdot)$ is bounded;
- in Young (1998), $f(\cdot)$ is unbounded but sublogarithmic;
- in Aghion and Howitt (1998, ch. 12.2), $f(\cdot)$ is assumed only to be increasing and concave.

These models are designed to yield endogenous growth, i.e. constant technology growth given constant research inputs. As a result, as can be seen from (3), these models do not match the finding that across many domains, ideas have gotten harder to find—sustained growth has required growing research inputs (Bloom et al., 2020)—or with the ways in which technological advances can allow growing inputs to be better parallelized.

In R&D function (2), by contrast, ideas may get harder to find ($\beta > 0$), so sustained growth may require growing research inputs. The prevailing technology limits the

[1] See Trammell (2025) for an overview.

extent to which many research inputs can be used productively, so a sudden jump in these inputs would not produce arbitrarily rapid technological progress. But as technology advances, so does our ability to absorb the growing research inputs need to sustain technological development.

On the AI front, this paper is closely related to existing work on the conditions under which automating R&D, or AI R&D in particular, will yield "explosive"—highly superexponential—growth in intelligence, technology, or output. Aghion et al. (2019), Davidson (2023), Erdil et al. (2025), Lifland et al. (2025), Davidson et al. (2026), and Jones and Tonetti (2026) all find a relevant "explosive growth" threshold in models of a similar spirit. The Davidson et al. (2026) model is especially similar to the generalized, multi-sector version of the model presented here.

This paper shows that the results of the papers above depend on their assumption that the elasticity λ of the speed of technological development $\dot{A}$ to research inputs R is constant, as in (1). When the parallelizability of R&D is modeled more realistically, as in (2), the conditions for superexponential growth are slightly narrower, and any transition to a superexponential regime induced by automating R&D may unfold much more gradually.

2 One-sector model

2.1 The conventional R&D function

Recall the conventional semi-endogenous R&D function:

$$g_{At} = \theta A_t^{-\beta} R_t^{\lambda}, \qquad \theta > 0,\ \beta > 0,\ \lambda \in (0,1) \tag{4}$$

where A denotes technology and R denotes research inputs. Functional form (4) captures the stylized facts that

1. ideas get harder to find ($\beta > 0$) and
2. R&D is imperfectly parallelizable ($\lambda < 1$).

More subtly, it also explains the fact that in many domains, we observe

3. constant $g_R > 0$ yields constant $g_A > 0$.

Given constant g_R, (4) is constant at the steady state

$$g_A^* = \frac{\lambda}{\beta} g_R. \tag{5}$$

If we observe constant g_R and g_A over a long period, we can identify $\lambda/\beta = g_A^*/g_R$.

2.2 An R&D function with bounded parallelizability

Facts 1–3 can also be matched by an R&D function in which, at any state of technology, our ability to parallelize R&D is bounded. We will work with a minimal deviation from (4):

$$g_{At} = \theta A_t^{-\beta} m\big(A_t^{\gamma}, R_t^{\lambda}\big), \qquad \theta > 0,\ \beta > 0,\ \ \lambda > 0,\ \ \gamma \geq \beta. \tag{6}$$

It would perfectly capture the spirit of the point I mean to make to posit that $m(\cdot)$ denotes a CRS, CES function with elasticity of substitution less than 1:

$$m\big(A^{\gamma}, R^{\lambda}\big) = \big(A^{\gamma\rho} + R^{\lambda\rho}\big)^{\frac{1}{\rho}}, \quad \rho < 0.$$

But up to an unimportant constant, none of the results change, and some can be expressed or proven more straightforwardly, when we take $m(\cdot)$ to denote the minimum. I will do so for simplicity.[2]

Note that the R&D function may equivalently be written

$$g_{At} = \theta A_t^{-\beta} m\big(A_t^{\gamma/\lambda},\ R_t\big)^{\lambda}. \tag{7}$$

This expression lends itself to a natural interpretation: the returns to scale to research inputs are still λ, insofar as they can be used at all, but the prevailing technology A only allows $A^{\gamma/\lambda}$ research inputs to be used productively at once.

Since $m(\cdot)$ bounds the parallelizability of R&D, imposing $\lambda < 1$ is no longer needed

[2] In fact, we could use any $m(\cdot)$ satisfying

i) CRS,
ii) continuity,
iii) monotonicity in both arguments from $m(x, 0) = m(0, y) = 0$, and
iv) the limits $m(x, \infty) = x$ and $m(\infty, y) = y$.

to maintain Fact 2. The condition

$$\gamma \geq \beta \tag{8}$$

is necessary for exponential technology growth to be possible: if $\gamma < \beta$, then even with $R = \infty$, g_A falls as A grows.

If inequality (8) is strict, technology growth is always constrained in the long run by a lack of research inputs, as in the conventional setting. If (8) holds with equality, however, long-run technology growth may fall into one of two regimes: one constrained by research inputs or, if research inputs grow quickly enough, one constrained by a lack of parallelization-facilitating technology.

Proposition 1 (Technology growth given constant research input growth)**.** *Given constant $g_R > 0$, the technology growth rate tends to*

$$g_A^* = \begin{cases} \frac{\lambda}{\beta} g_R, & \gamma > \beta; \\ \min\left(\frac{\lambda}{\beta} g_R, \theta\right), & \gamma = \beta. \end{cases}$$

Proof. See Appendix A.1. □

The $\gamma = \beta$ case is knife-edge, or at least arguably so (see footnote 5 and the preceding discussion below). Nevertheless, it illustrates in a simple way how parallelizability constraints may begin to bind if research inputs start growing more quickly. Past some point, doubling the growth rate of research inputs may fail to double the technology growth rate. The importance of a hard parallelizability constraint will arise more pervasively when we consider the implications of automating R&D, and still more pervasively when we consider the multi-sector model.

Note that in practice we can still expect (5) to hold in general, and thus infer $\lambda/\beta = g_A^*/g_R$ from long-term technology and research input growth rates, even in the $\gamma = \beta$ case. Growing R so quickly that $\frac{\lambda}{\beta} g_R > \theta$ would be a waste of inputs.

Analogy to the case of fixed R&D inputs. Observe that R&D function (6) could apply just as well to the case in which there are two inputs to the research process, of which one (R) is accumulable, and the other—a natural resource, say—is in fixed supply, its effective supply growing only as technological progress lets us make do with less of it (with A^{γ}). This paper could therefore be reframed as introducing an R&D-based growth model in which R&D depends on some inputs in fixed supply.

This reframing is especially relevant to the case of automated AI R&D. In Section 4, we consider the implications of automating AI R&D under not only parallelizability constraints but also hard constraints on the quantity of compute and data. The constraints, at least as modeled here, will be essentially equivalent.

2.3 Implications for automating R&D

Suppose R&D is automated, so that it can now be performed entirely by capital. We can model this shift in two ways.

"Infinite" research inputs. Suppose the shift happens suddenly, such that for all practical purposes (at least for some time) research inputs become infinitely abundant. This scenario would correspond, for instance, to one in which

i) A is an index of AI capabilities;

ii) fixing today's data and hardware, new insight from human AI researchers can serve as a gross substitute for all other inputs to AI development; and

iii) we introduce the ability to simulate an AI researcher efficiently, so that at once we can simulate billions of (AI researcher) "geniuses in a data center".[3]

The implications for AI progress after jumping to $R = \infty$ depend greatly on the R&D function.

On the conventional approach (4), we have

$$g_{A_t} = \theta A_t^{-\beta} \cdot \infty = \infty \, ,$$

and A jumps arbitrarily high in no time.

But in practice, even if the research inputs were infinite, the technology available at a given time would put constraints on the extent to which the inputs could be absorbed productively. Firms and labs cannot instantaneously multiply their research output suddenly by hiring all and sundry; while they build their research teams, they must

[3] To quote Amodei (2024).

develop the infrastructure to manage and coordinate their work.[4,5] Incorporating this point, via an R&D function in which the [here, AI] technology limits parallelizability (6), we have

$$g_{A_t} = \theta A_t^{-\beta} m\left(A_t^{\gamma}, \infty\right) = \theta A_t^{\gamma-\beta}.$$

Once the constraint imposed by a lack of research inputs has been relieved, technology advances more quickly, but not instantaneously. If γ is low, so that relieving parallelizability constraints is difficult, technology may take a long time to explode or may not explode at all.

Proposition 2 (Technology growth given infinite research inputs)**.**
Given infinite R,

- *If $\gamma = \beta$, the technology growth rate jumps permanently to θ.*
- *If $\gamma > \beta$, technology grows hyperbolically to a vertical asymptote at*

$$t^* = \frac{A_0^{-(\gamma-\beta)}}{\theta(\gamma - \beta)}.$$

In the $\gamma = \beta$ case, the growth rate θ is only weakly higher than the steady-state technology growth rate given constant g_R, namely $\min\left(\frac{\lambda}{\beta} g_R, \theta\right)$.

Likewise, in the $\gamma > \beta$ case, low $\gamma - \beta$ may render the vertical asymptote be arbitrarily distant.

Research inputs proportional to technology. Suppose that immediately after R&D automation, labor is no longer employed directly in R&D, but the quantity of capital employed in R&D ("research capital") remains meaningfully finite. If the labor

[4] I am hardly alone in having been turned down when offering free research assistance, and in having turned down the offer of a free research assistant!

[5] Incidentally, this example suggests a reason why the purely parallelization-technology-bounded growth process might sometimes be exponential, i.e. why the $\gamma = \beta$ case might not be knife-edge, at least in some domains: each existing employee can only hire and properly integrate so many per unit time. See e.g. Garicano and Rossi-Hansberg (2012) for more on this sort of friction to expansion. That said, technological advances can also allow for faster "onboarding"; AIs that do everything ever more quickly would be able to integrate new "employees" ever more quickly as well.

producing this research capital is fixed and normalized to 1, and A denotes the productivity of this labor, automating R&D yields $R_t = A_t$.[6] Likewise, $R_t = A_t$ if the stock of research capital is itself fixed at 1 but A denotes its own productivity: e.g. the number of virtual researchers that can be run on fixed computer hardware.

Under these scenarios, the conventional R&D function (4) predicts

$$g_{At} = \theta A_t^{\lambda-\beta}. \tag{9}$$

Automating R&D thus delivers hyperbolic technology growth on fixed labor or hardware if and only if $\lambda/\beta > 1$.[7] Recall also from (5) that λ/β can be identified from the input and technology growth rates observed in the pre-automation era by

$$\frac{\lambda}{\beta} = \frac{g_A^*}{g_R}. \tag{10}$$

Motivated by these observations, Eth and Davidson (2025) compare the growth rates of various measures of AI quality (g_A) to the growth rates of the number of relevant AI researchers and other inputs (g_R). Finding $g_A > g_R$ in many domains, they conclude that automating AI R&D would likely yield the "intelligence explosion"—hyperbolic growth in AI quality—characterized by (9) in the $\lambda/\beta > 1$ case.

The bounded-parallelizability R&D function (6) instead predicts

$$\begin{aligned} g_{At} &= \theta A_t^{-\beta} m\left(A_t^{\gamma}, A_t^{\lambda}\right) \\ &= \theta A_t^{\kappa}, \quad \kappa \equiv \min\left(\gamma - \beta, \lambda - \beta\right). \end{aligned}$$

As usual, we can infer λ/β from historical g_A^*/g_R (see Prop. 1 and the following discussion). Furthermore, if $\lambda/\beta < 1$, we can still infer that automating R&D will not deliver even exponential technology growth (on fixed labor or hardware): $\lambda - \beta < 0$. And if $\lambda/\beta = 1$, we can still infer that automating R&D will deliver precisely exponential technology growth at rate θ, since $\lambda - \beta = 0$ in this case, and $\gamma - \beta \geq 0$ by (8). If $\lambda/\beta > 1$, however, we can no longer infer an explosion, let alone a sudden explosion, in technology. What follows depends on γ.

[6] Assuming for simplicity that research capital depreciates instantaneously.

[7] And exponential growth at rate θ on fixed labor or hardware if $\lambda/\beta = 1$. If $\lambda/\beta < 1$, exponential technology growth continues to require an exogenous source of exponentially growing research inputs.

Proposition 3 (Technology growth given research inputs proportional to tech)**.** *Given $R \propto A$: (i) If $\lambda/\beta < 1$, g_{At} falls over time. (ii) If $\lambda/\beta = 1$, $g_{At} = \theta$. (iii) If $\lambda/\beta > 1$,*

- *If $\gamma = \beta$, $g_{At} = \theta$.*
- *If $\gamma \in (\beta, \lambda)$, technology growth is hyperbolic, but slower than in (9).*
- *If $\gamma \geq \lambda$, technology growth is hyperbolic as in (9).*

In effect, $\gamma - \lambda$ denotes the ease of relieving parallelizability constraints *relative to that of growing the stock of research inputs.* Above the growth path traced by the historical exponential growth in research inputs R, there has hung a "parallelizability ceiling" imposed by $A^{\gamma-\beta}$. This ceiling must have risen at least exponentially as well, or else with enough time we would have hit it: hence the assumption of $\gamma \geq \beta$. But if automating R&D would allow effective research inputs to grow superexponentially, we do hit the growth ceiling imposed by our ability to use these inputs in parallel, unless—because $\gamma \geq \lambda$—the ceiling itself rises at least as quickly.

This will allow us to estimate γ, once we introduce the multi-sector model, from the returns to R&D on parallelizability-enhancing technologies.

3 Technology network model

Often, the technology that facilitates progress on one technological dimension, A_i, is not A_i itself but technology of another sort, A_j. Software development is sped in part by faster computer hardware, for instance, not just better software. Davidson et al. (2026) study the implications of these cross-sector technology spillovers under the standard assumption that the contribution of each technology to each other technology's R&D is Cobb-Douglas: that for a **network** of n technologies,

$$g_{A_i} = \theta_i R_i^{\lambda} \, \Pi_{j=1}^{n} A_j^{\phi_{ij}} \tag{11}$$

(dropping time subscripts). It is natural to assume that $\phi_{ii} < 0$, since ideas get harder to find in any particular domain, and that $\phi_{ij} \geq 0$ for $i \neq j$.

By the same token, the technological advances *required* to speed research on technology A_i may be advances in other technologies. Advances in AI capabilities, for instance, have largely been driven by increases in the compute used in training; but even if compute were the only input to AI research, increases in training compute would not speed AI progress indefinitely, because our hardware architecture does not allow us to parallelize training runs perfectly (Erdil and Schneider-Joseph, 2024). If today's technology only lets us make effective use of 10^{23} FLOP/s in a given training run, a 10^{30}-FLOP training run cannot take less than 10^7 seconds (3–4 months), even if we have infinite hardware. As a result, continued acceleration in AI capabilities may eventually require hardware improvements, not just be facilitated by them.

To model this phenomenon flexibly, we will suppose that the development of technology A_i is a gross-complements combination of (i) research inputs R_i and (ii) up to $Z \geq 1$ technological constraints on the extent to which these inputs can be put to good use simultaneously. Relieving each of these constraints may require advances in a single technology, or multiple technologies may each contribute in a Cobb-Douglas fashion. Technologies may also increase effective research inputs in a Cobb-Douglas fashion. So, for each i, we have

$$g_{A_i} = m\left(R_i^{\lambda_i} \cdot \theta_{i0}\Pi_{j=1}^{n} A_j^{\phi_{ij0}},\ \theta_{i1}\Pi_{j=1}^{n} A_j^{\phi_{ij1}},\ \dots,\ \theta_{iZ}\Pi_{j=1}^{n} A_j^{\phi_{ijZ}}\right), \tag{12}$$

where $\phi_{iiz} < 0$ and $\phi_{ijz} \geq 0$, $j \neq i$ for each $z = 0, 1, ..., Z$.

R&D function (12) reduces to the R&D function (6) of Section 2, in the $Z = n = 1$ case, since the $\theta A^{-\beta}$ of (6) may be distributed across the CRS function $m(\cdot)$. Likewise, (12) reduces to the R&D function (11) from Davidson et al. (2026) in the $Z = 0$ case, for any n, though note that we here assume $Z \geq 1$. Also, even with $Z \geq 0$, (12) can accommodate the possibility that g_{A_i} exhibits a constant elasticity to A_j, unconstrained by any bottlenecks (as in (11)), because an $A_j^{\phi_{ij}}$ term can be distributed across $m(\cdot)$.

If technologies face different numbers of constraints, Z can denote the maximum number of constraints across the n technologies. Then for a technology with $\underline{Z} < Z$ constraints, entries $\underline{Z} + 1, ..., Z$ may be duplications of earlier entries, without loss of generality.

3.1 Semi-endogenous tech growth before automated R&D

Let Φ_z denote the $n \times n$ matrix with $[\Phi_z]_{ij} = \phi_{ijz}$. Observe that Φ_z is negative on the diagonal and nonnegative off the diagonal.

Let $D \equiv -\text{diag}(\phi_{1,1,0}, ..., \phi_{nn0})$ denote the diagonal matrix whose diagonal is the negative of Φ_0's diagonal. Assume that

$$\rho\left(D^{-1}(\Phi_0 + D)\right) < 1,$$

where $\rho(\cdot)$ denotes the spectral radius, i.e. the largest eigenvalue. This generalizes the $\beta > 0$ condition in the conventional R&D function (4): it is equivalent to the condition that growth is not explosive when research inputs are constant or grow exponentially (see Davidson et al. (2026)).

Given that the research inputs grow at constant rates $g_R \equiv \left(g_{R_1}, ..., g_{R_n}\right) \gg 0$, we will now identify a condition under which each technology also grows exponentially in the limit. That is, the condition will maintain that for each i, g_{A_i} converges to the positive, steady-state value of i's *input constraint*—"constraint 0"; the constraint imposed by a lack of research inputs—so that for large t

$$g_{A_i t} = R_{it}^{\lambda_i} \cdot \theta_{i0} \Pi_{j=1}^{n} A_{jt}^{\phi_{ij0}}, \tag{13}$$

and no A_i is constrained to subexponential growth due to an intrinsic limit on the rate of growth in the quantity of effective research inputs that can be absorbed. This generalizes the $\gamma \geq \beta$ condition of (6). We will then find the steady-state technology growth rates $g_A^* \equiv \left(g_{A_1}^*, ..., g_{A_n}^*\right) \gg 0$.

Proposition 4 (Multi-sector tech growth given constant research input growth). *Given constant $g_R \gg 0$, let Λ denote the n-vector with $\Lambda_i = -\lambda_i g_{R_i}$, and let $\bar{g}_A \equiv \Phi_0^{-1}\Lambda$.*

1. *If $\Phi_z \bar{g}_A \gg 0 \ \ \forall z \geq 1$, then g_A satisfies (13) for large t.*
2. *If g_A satisfies (13) for large t, then $g_A^* = \bar{g}_A$.*

Proof. For part 1, see Appendix A.2. For part 2, see Davidson et al. (2026), Prop. 1.1. □

For intuition, observe that $\bar{g}_A$ denotes the steady-state growth rate vector that obtains when tech growth is input-constrained. It is analogous to $\frac{\lambda}{\beta} g_R$ in the single-sector setting: Λ is analogous to $-\lambda g_R$, and Φ_0^{-1} is analogous to $-1/\beta$. In the $n = 1$ case, $\phi_{1,1,0}$ is precisely the quantity denoted $-\beta$ in Section 2.

Likewise, Φ_z is analogous to $\gamma - \beta$: in the $n = 1$, $Z = 1$ case, $\phi_{1,1,1}$ is precisely the quantity denoted $\gamma - \beta$ in Section 2. The $\Phi_z \bar{g}_A \gg 0$ condition is then analogous to the condition that $(\gamma - \beta)\frac{\lambda}{\beta} g_R > 0$, or equivalently that $\gamma > \beta$. It implies that when the input constraints are constant, and g_A is in steady state, the parallelizability constraints grow at a positive rate and so are eventually non-binding.

For simplicity, we will not cover the edge case in which some or all entries of $\Phi_z \bar{g}_A$ are zero. In this case, whether growth is input- or parallelizability-constrained depends on a comparison between $\bar{g}_A$ and the matrix of θ_{ij} coefficients, analogous to the point that $g_A^* = \min\left(\frac{\lambda}{\beta} g_R, \theta\right)$ in the $\gamma = \beta$ case of Proposition 1.

3.2 Technology growth given automated R&D

In Section 2.3, we considered two ways of modeling automated R&D: as an elimination of the input constraint, and as a shift to a regime in which the quantity of research inputs is proportional to the technology level. In the multi-sector model of this section, these two modeling approaches are equivalent.

- On replacing R_i with A_j for some j, expression (12) for g_{A_i} reduces to

  $$g_{A_i} = m\left(\theta_{i0} \Pi_{j=1}^n A_j^{\phi_{ij0}}, \ldots, \theta_{iZ} \Pi_{j=1}^n A_j^{\phi_{ijZ}}\right), \tag{14}$$

  for some new exponent matrices Φ_z which are, again, non-negative off the diagonal and, now only for $z \geq 1$, negative on the diagonal. Replacing R_i with any Cobb-Douglas combination of technologies, or any nesting of Cobb-Douglas and minimum functions as in (14), likewise yields an expression for g_{A_i} that reduces to the above.

- Eliminating the input constraint yields (14) as well, except that the constraint indexing now begins at 1 rather than 0:

  $$g_{A_i} = m\left(\theta_{i1} \Pi_{j=1}^n A_j^{\phi_{ij1}}, \ldots, \theta_{iZ} \Pi_{j=1}^n A_j^{\phi_{ijZ}}\right). \tag{15}$$

We will work with the latter formulation, so that the minimum number of constraints after automating R&D, which obtains when $Z = 1$, is 1 rather than 2.

Simplifying assumptions and notation. We will consider the case in which R&D is automated for every technology under consideration, i.e. g_{A_i} takes form (15) for $i = 1, ..., n$. This is substantively without loss of generality: since we are considering the explosive technology growth that can follow from automated R&D, R_i for technologies i whose R&D is not automated can be assumed to be roughly constant and incorporated into θ_{i0}.

Our technology network exhibits **strict complementarity** if the off-diagonal elements of each Φ_z are strictly positive. This condition ensures that, if $\lim_{t\to T} A_{it} = \infty$ for some i, then for all j we have

- $\lim_{t\to T} g_{A_j t} = \infty$ (though not necessarily $\lim_{t\to T} A_{jt} = \infty$) and
- A_{jt} undefined for $t > T$,

rather than a scenario in which some variables grow hyperbolically to T while others grow (sub-)exponentially, or grow hyperbolically to vertical asymptotes at later dates.[8] A technology network exhibiting strict complementarity **reaches a singularity** at T if $\lim_{t\to T} A_{it} = \infty$ for some i.

Let $\mathcal{Z} \equiv \{1, ..., Z\}$ and $\mathcal{N} \equiv \{1, ..., N\}$, so that $\mathcal{Z}^{\mathcal{N}}$ denotes the set of functions from technologies to constraints. Then $\zeta \in \mathcal{Z}^{\mathcal{N}}$ denotes the technology network in which technology i faces only the single constraint $\zeta_i \in \mathcal{Z}$. Observe that if our original technology network exhibits strict complementarity and reaches a singularity, then so does ζ. Let T_ζ then denote the date at which ζ reaches a singularity. For a guide to estimating T_ζ numerically, for a given ζ, see Davidson et al. (2026).

Proposition 5 (Multi-sector tech growth given automated R&D).
Suppose that R&D throughout a technology network is automated, so that the growth of each technology i takes form (15).

[8] If we define $A_{it} = \infty$ for $t \geq T$ (allowing $g_{A_i t}$ to be undefined when $A_{it} = \infty$), and stipulate that A_{jt} is defined and non-decreasing for all j, t, it follows that $A_{jt} = \infty$ for $t > T$, for all j. Proof: if $A_{jt'} < \infty$ for some $t' > T$, then $A_{jt'} < \infty$ for $t \in (T, t')$ by monotonicity. Then $A_{it} = \infty$ yields $g_{A_j t} = \infty$ for $t \in (T, t')$, contradicting the finitude of $A_{jt'}$.

1. *If*

   a. *the technology network exhibits strict complementarity and*

   b. *for some* $g \gg 0$ *we have* $\Phi_z g \gg 0\ \forall z$,

   the technology network reaches a singularity at some date T.

2. $T \geq \max_{\zeta \in \mathcal{Z}^{\mathcal{N}}} T_\zeta$ *(but has no closed-form solution).*

Proof. See Appendix A.3. □

Note that condition 1b weakens the condition imposed in Proposition 4 to ensure that exponential technology growth is possible with inputs growing at rates g_R.

Proposition 5 generalizes the message from Proposition 3 in the single-sector case: though parallelizability constraints not strong enough to block exponential growth generically also do not block singularities (5.1), they may delay them (5.2). Here, we see that the larger $\mathcal{Z}^{\mathcal{N}}$ is—the larger the technology network and the greater the number of technological requirements to developing each technology—the more opportunities there are for parallelizability constraints to delay the technological explosion.

4 Application to AI R&D

We will now apply the model of Section 3, of a technology network with parallelizability constraints, to the case of AI R&D. This is the application that motivates the paper: AI R&D may soon be automated (see most recently Favaro and Clark (2026))—AI may soon be able to *recursively self-improve*—and it is valuable to understand how quickly advances in AI will subsequently proceed.

This modeling exercise is inevitably somewhat speculative and stylized. That said, scaling laws make it possible to write down the production function for advances in AI with greater granularity, in many ways, than the production function for advances in other domains of technology.

4.1 The basics of recursive self-improvement

Suppose we soon develop an AI model capable, on one of today's data centers, of autonomously carrying on the work of a frontier AI research team. The first question in the study of AI's recursive self-improvement is whether AI performance will then explode or fizzle, holding fixed the data and computer hardware ("compute") the model has to work with.

The answer is non-obvious because AI advances to date have been driven by simultaneous improvements in data, compute, and the algorithmic insights supplied by AI researchers. This makes it difficult to determine whether better algorithms alone, as an automated AI research team could develop without waiting for new data or compute to accumulate, would have sufficed (Gundlach et al., 2025). Furthermore, even if algorithmic progress does suffice to drive AI progress more or less indefinitely, using the stocks of data and compute available in a frontier data center today, it may proceed slowly. Intuitively, after each doubling of algorithm quality, further algorithmic improvements may get more than twice as difficult to develop, in which case AI advances on fixed compute and data will decelerate with time.

If we conclude that the recursive self-improvement following automated AI R&D would suffice for not only large but ever-accelerating advances in AI capabilities—an *intelligence explosion*—the second question is how quickly the explosion will take place.

Simplest model: labor-only. To illustrate the basics of recursive self-improvement as simply as possible, begin with the standard semi-endogenous growth model. Letting A denote some index of *algorithm quality*,

$$g_{At} \propto A_t^{-\beta} R_t^{\lambda}.$$

Before AI R&D has been automated, say $R_t^{\text{pre}} = L_t$, where L_t denotes human research labor. For now, abstract from the need for data and compute in AI training entirely.

Once AI R&D is fully automatable, suppose that it is fully automated in practice, so that the research inputs R_t^{post} are proportional to the stock of research capital $\bar{K}$ (composed primarily of data and compute). To assess whether this can lead to an explosion in AI progress, at a rate not bottlenecked by capital accumulation, we assume that $\bar{K}$ is held fixed. As algorithms improve, however, ever less capital is needed to do the work of one human researcher. If each doubling of algorithm quality (in the sense characterizing A) comes with a doubling in the number of human researcher-equivalents that can be run on fixed compute, then effective research inputs after the automation of AI R&D equal $A_t\bar{K}$, so that

$$g_{At} \propto A_t^{-\beta}\left(A_t\bar{K}\right)^{\lambda} \propto A_t^{\lambda-\beta}. \tag{16}$$

Algorithm quality A—or equivalently, here, the number of human-researcher-equivalents that can be run on a unit of compute—then grows subexponentially if $\lambda < \beta$, exponentially if $\lambda = \beta$, and superexponentially if $\lambda > \beta$. In particular, if $\lambda > \beta$, A grows hyperbolically, to a singularity whose date is proportional to $1/(\lambda - \beta)$: a result precisely analogous to that of Proposition 2.

Adding capital. Now suppose that, before AI R&D is automated, the research input is a Cobb-Douglas combination of labor and capital. Because R is raised to a power in the R&D function, we can write $R_t^{\text{pre}} = L_t\bar{K}^{\xi}$ without loss of generality, so that the exponent on labor remains λ. Then the lessons of the labor-only model are precisely maintained. When AI R&D is automated, so that a quantity proportional to $A\bar{K}$ takes the place of L in R^{post}, the R&D function reduces to (16).

Suppose instead, however, that the cognitive research work currently done by labor, and the training and experiments executed by capital, are gross complements in AI R&D. This assumption is arguably more intuitive, and it has some empirical support from Whitfill and Wu (2025). Then the implications of automating AI R&D depend crucially on whether algorithmic improvements augment the training and experiments as much as the cognition. In particular, R&D inputs after AI R&D automation in this case equal

$$R_t^{\text{post}} = m\left(sA_t\bar{K}, (1-s)A_t^{\gamma}\bar{K}\right),$$

where s denotes the fraction of the capital used for cognition—i.e. inference that replaces human research work—as opposed to experimentation. If $\gamma \geq 1$, cognition is the long-run bottleneck, and the R&D function again ultimately reduces to (16). Otherwise, training and experiment compute is the long-run bottleneck, and the R&D function reduces to

$$g_{At} \propto A_t^{-\beta}\left((A_t\bar{K})^{\gamma}\right)^{\lambda} \propto A_t^{\gamma\lambda-\beta}.$$

This is structurally the same model, but with a parameter change: the existence and date of a singularity depend on $\gamma\lambda - \beta$, rather than $\lambda - \beta$.

As noted at the end of Section 2, this is precisely analogous to the point that effective research inputs or parallelization technology may be the long-run bottleneck to R&D in a given domain, depending on the exponents.

More labor vs. smarter labor. None of the models above incorporate the fact that, when a wave of algorithmic improvements lets us simulate twice as many AI researchers on fixed capital (when "A" doubles), it will also typically let us simulate a smaller number of smarter, more compute-intensive AI researchers. The less numerous but smarter "research team" may sometimes make faster progress than a team of $A_t\bar{K}$ researchers at human-like capacity, so the rate of AI progress on fixed capital is only *lower-bounded* by, say, $A_t^{-\beta}(A_t\bar{K})^{\lambda}$.

A model in which the the "intelligence" of each simulated AI researcher grows alongside their number is roughly equivalent to one like the above but with a lower value of β. For more, see Lifland et al. (2025). For the remainder of this section, we will work with the simpler case in which increases in AI quantity and quality are equivalent.

4.2 AI R&D under parallelizability constraints

One might think that with enough chips of today's designs, we would be able to accelerate AI R&D indefinitely: even without designing or conducting any new experiments, we could run arbitrarily many training runs in parallel, test the outcome of each in parallel, and pick the one that scores best in testing. In the extreme, the argument goes, we could simply simultaneously test every possible weighting of, say, a 10-trillion-parameter model.

This is incorrect. Even if each parameter were only a single bit, there would be a whopping $2^{10\text{T}}$ 10-T-parameter models. Merely storing the weights of a 10-T-parameter model takes over 20TB, which using today's technology must occupy at least ~8cm^3.[9] Storing all possible 10-T-parameter models, using today's storage, would thus require a cube with sides over 10^{3005}km long, which would take light around 10^{2992} years to traverse.

Even infinite computation, therefore, would not deliver arbitrary advances quickly. AI R&D is subject to parallelizability constraints, like those in any other domain, and we will propose a parallelizability-constrained model of AI R&D accordingly.

Loss. The loss[10] a "pre-trained" AI model achieves on a testing dataset decreases roughly power-functionally in the size of the model—the number of parameters N—and the quantity of data D drawn from a similar distribution:

$$\mathcal{L} \approx \left(A_1 N\right)^{-\alpha_1} + \left(A_2 D\right)^{-\alpha_2} \tag{17}$$

(see especially Hoffmann et al. (2002)). $C = 6ND$ FLOPs are needed to train a model of size N on a dataset of size D.[11] An increase in training compute can thus lower loss by allowing the training run to absorb more data and/or by training a larger model. Loss can also fall even without increasing N or D (nor, as a result, training compute C) due to *algorithmic progress*: an increase in *parameter efficiency* A_1 or *sample efficiency* A_2. We will set aside advances in post-training.

We will suppose that a model exhibiting some sufficiently low loss $\mathcal{L}^*$ is capable of automating AI R&D: that is, is capable of autonomously developing increases to A_1 and A_2. As noted at the end of Section 4.1, once our algorithms, compute, and data have advanced to the point that $\mathcal{L}^*$ is achievable, we assume that AI R&D is carried on by an increasing number of instances of a model at loss $\mathcal{L}^*$, rather than models of ever lower loss.

By (17), effective parameters $A_1 N$ and effective data $A_2 D$ are gross complements in the production of any index of "performance" that is inversely related to loss $\mathcal{L}$. To maintain a target loss of $\mathcal{L}^*$, therefore, large increases in parameter efficiency must

[9] The densest storage available today appears to be the 246TB Kioxia LC9 E3.L SSD, occupying just over 81cm^3.

[10] Putting aside irreducible loss.

[11] Assuming one epoch per data point, and assuming that the model is dense.

approximately pay off in proportional decreases to model size, and large increases in sample efficiency must in proportional decreases in data usage.[12] We will assume for simplicity that this relationship is exact: i.e. that at algorithm qualities A_1, A_2, we train a model of size $N^* \propto 1/A_1$ on a dataset of size $D^* \propto 1/A_2$.

Algorithmic progress. Like Whitfill and Wu (2025), we will say that algorithmic advances of either kind require both cognition and experiments. For $i = 1, 2$,

$$g_{A_i} = A_i^{-\beta_i} m(\text{Xpm}_i, \text{Cog}_i), \tag{18}$$

where Cog_i is an index of the cognitive throughput designing experiments to improve algorithms on dimension i, and Xpm_i indexes the throughput of experiments—and the full training runs that take advantage of new well-performing algorithms. I will refer to the latter together simply as "experiments".

I conflate experiments and training runs because, following Whitfill and Wu (2025), we can arguably model experiments as partial training runs. The compute requirement of an experiment or training run at technology state A_1, A_2 is thus proportional to $N^* D^*$ and thus in turn to $1/(A_1 A_2)$.

Effective experiments. If each experiment or training run could be fully parallelized, a fixed compute stock of some size could support $x \propto A_1 A_2$ experiments or training runs per unit time. Then, with $\sigma_{\text{Xpm},i}$ denoting the fraction of our compute allocated to experiments of type i, we could straightforwardly write $\text{Xpm}_i = \theta_{i1} \cdot \sigma_{\text{Xpm},i} A_1 A_2$.

In fact, however, training a model of size N on a dataset of size D requires a number of serial computations proportional to $D \times d$, where d is the *depth* of the model, which scales roughly with $N^{1/3}$.[13] On a given stock of compute allocated to experiments, therefore, we can run no fewer than $\propto A_1^{2/3}$ streams of experiments in parallel, with each stream completing $\propto A_1^{1/3} A_2$ experiments per period. Modeling the diminishing

[12] In particular, achieving loss $\mathcal{L}^*$ requires $N \gtrsim \mathcal{L}^{*-\frac{1}{\alpha_1}} / A_1$ and $D \gtrsim \mathcal{L}^{*-\frac{1}{\alpha_2}} / A_2$.

[13] The exponent is precisely 1/3 if the scale of the model does not affect its relative dimensions. Erdil and Schneider-Joseph (2024) conclude that larger models can be made relatively less deep, so that the exponent is 0.27. We will use 1/3 for simplicity.

returns to simultaneous experiments in the usual way, and fixing $\sigma_{\text{Xpm},i}$, we have

$$\text{Xpm}_i = \theta_{i1} \cdot \left(A_1^{\frac{1}{3}} A_2\right) A_1^{\frac{2}{3}\lambda_i} = \theta_{i2}\, A_1^{\frac{1+2\lambda_i}{3}} A_2. \tag{19}$$

Effective cognition. The number of FLOPs a model needs to generate an inference token is proportional to the model's size.[14] As with the experiments, each inference token requires serial computations scaling roughly with $N^{1/3}$, so that in technology state A_1, A_2, we cannot have a single "virtual researcher" producing experiment ideas arbitrarily quickly, but must make do with $\propto A_1^{2/3}$ "researchers" each producing $\propto A_1^{1/3}$ tokens per unit time.

As with the robot engineers of Section 1, coordinating the work of a growing population of researchers requires a third kind of technology A_3, which we might call *agent parallelizability*. This technology indexes the number of agents that can productively be used at once to generate valuable experiment plans. Anthropic's recent development of "agent teams" in Opus 4.6, for example,[15] was an increase in agent parallelizability.

Thus

$$\text{Cog}_i = A_1^{\frac{1}{3}} m\left(\theta_{i2} A_1^{\frac{2}{3}}, \theta_{i3} A_3\right) \tag{20}$$

Since advances in agent parallelizability too require cognition and experiments, we will say that g_{A_3} takes form (18), with Cog_3 and Xpm_3 given by (19)–(20).

The technology network. In sum, from (18)–(20), the technology network for AI R&D in this stylized setting can be characterized by

$$\begin{aligned}
g_{A_1} &= m\Big(\theta_{1,1} A_1^{\frac{1+2\lambda_1}{3}-\beta_1} A_2, \quad \theta_{1,2} A_1^{1-\beta_1}, \quad \theta_{1,3} A_1^{\frac{1}{3}-\beta_1} A_3 \quad \Big),\\
g_{A_2} &= m\Big(\theta_{2,1} A_1^{\frac{1+2\lambda_2}{3}} A_2^{1-\beta_2}, \quad \theta_{2,2} A_1 A_2^{-\beta_2}, \; \theta_{2,3} A_1^{\frac{1}{3}} A_2^{-\beta_2} A_3\Big),\\
g_{A_3} &= m\Big(\theta_{3,1} A_1^{\frac{1+2\lambda_3}{3}} A_2 A_3^{-\beta_3}, \; \theta_{3,2} A_1 A_3^{-\beta_3}, \; \theta_{3,3} A_1^{\frac{1}{3}-\beta} A_3^{1-\beta_3}\Big).
\end{aligned}$$

[14] In particular, a model of size N takes $2N$ FLOPs to produce a token.

[15] See Anthropic (2026).

If we know $\{\lambda_i\}$ and $\{\beta_i\}$, Proposition 5 can then be used to test whether the network exhibits a singularity and, if so, to lower-bound its date.

5 Conclusion

This paper explores a hitherto overlooked class of bottlenecks to R&D: hard parallelizability bottlenecks.

These bottlenecks do feature prominently in the informal conversation over whether AI will be able to deliver a radical acceleration to technological progress. Some argue that even in a world of extremely advanced AI, progress in important domains, such as medicine, will be bottlenecked by a need for serial experiments (see e.g. Teslo (2026)). Others reply that sufficiently advanced AI could compress research timelines across the board, e.g. because with enough compute, we will be able to simulate the human body in such detail that slow, physical experimentation is unnecessary (see e.g. Kurzweil (2024)). This paper offers a model capturing both insights. Today, we do not have the technology that would let us to absorb large quantities of compute in ways that radically accelerate medical research. In the future, we may. But developing this technology itself will take time.

As modeled here, parallelizability bottlenecks are analogous to R&D inputs whose supply is fixed but whose effective supply may grow as technology advances. The model is therefore relevant to any domain in which R&D depends on an input in fixed supply. The application to parallelizability may be read as a case for thinking that such inputs are more prevalent than sometimes appreciated.

Whether parallelizability constraints will meaningfully slow AI R&D in particular, once it has been automated, depends on the parameters in a technology network like that outlined at the end of Section 4. Many of these parameters have not yet been estimated, but I hope this exercise motivates efforts to estimate them.

References

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–51.

_ **and** _ , *Endogenous Growth Theory*, Cambridge, MA: MIT Press, 1998.

__ , **Antonin Bergeaud, Timo Boppart, and Jean-Félix Brouillette**, "Resetting the Innovation Clock: Endogenous Growth through Technological Turnover," 2025. Working paper.

__ , **Benjamin F. Jones, and Charles I. Jones**, "Artificial Intelligence and Economic Growth," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, 2019, pp. 237–82.

Amodei, Dario, "Machines of Loving Grace," 2024. Blog post.

Anthropic, "Introducing Claude Opus 4.6," 2026.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, 2020, *110* (4), 1104–44.

Davidson, Tom, "What a Compute-Centric Framework Says About Takeoff Speeds," Technical Report, Coefficient Giving 2023.

__ , **Basil Halperin, Thomas Houlden, and Anton Korinek**, "When Does Automating AI Research Produce Explosive Growth? Feedback Loops in Innovation Networks," 2026. Working paper.

Erdil, Ege and David Schneider-Joseph, "Data movement limits to frontier model training," Technical Report, Epoch AI 2024.

__ , **Andrei Potlogea, Tamay Besiroglu, Anson Ho, Jaime Sevilla, and Matthew Barnett**, "GATE: An Integrated Assessment Model for AI Automation," Technical Report, Epoch AI 2025.

Eth, Daniel and Tom Davidson, "Will AI R&D Automation Cause a Software Intelligence Explosion?," Technical Report, Forethought 2025.

Favaro, Marina and Jack Clark, "When AI builds itself," Technical Report, Anthropic Institute 2026.

Garicano, Luis and Esteban Rossi-Hansberg, "Organizing Growth," *Journal of Economic Theory*, 2012, *147* (2), 623–656.

Gundlach, Hans, Alex Fogelson, Jayson Lynch, Ana Trisovic, Jonathan Rosenfeld, Anmol Sandhu, and Neil Thompson, "On the Origin of Algorithmic Progress in AI," 2025.

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre, "Training Compute-Optimal Large Language Models," in S. Koyejo, S. Mohamed,

A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems*, Vol. 35 2002, pp. 30016–30030.

Jones, Charles I., "R&D-based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 495–525.

_ **and Chris Tonetti**, "Past Automation and Future A.I.: How Weak Links Tame the Growth Explosion," 2026. Working paper.

Kurzweil, Ray, "How AI Supercharges Drug Research," *Science Friday*, 2024.

Lashkari, Danial, "Innovation Policy in Theories of Knowledge Diffusion and Selection," 2023. Working paper.

Lifland, Eli, Brendan Halstead, Alex Kastner, and Daniel Kokotajlo, "AI Futures Model," Technical Report, AI Futures Project 2025.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), 71–102.

Teslo, Ruxandra, "AI Will Not Solve Clinical Trials," *Asimov Press*, 2026.

Trammell, Philip, "The Extreme Inefficiency of Expanding Variety in Endogenous Growth Theory," 2025. Working paper.

Whitfill, Parker and Cheryl Wu, "Will Compute Bottlenecks Prevent an Intelligence Explosion?," 2025.

Young, Alwyn, "Growth without Scale Effects," *Journal of Political Economy*, 1998, *106* (1), 41–63.

A Proofs

A.1 Proof of Proposition 1

Let

$$q_t \equiv \frac{A_t^{\beta}}{R_t^{\lambda}}.$$

When the input constraint binds, we have $m\left(A_t^{\gamma}, R_t^{\lambda}\right) = R_t^{\lambda}$, so

$$\begin{aligned} & g_{At} = \theta / q_t \\ \implies \; & g_{qt} = \beta g_{At} - \lambda g_R = \beta\theta / q_t - \lambda g_R. \end{aligned}$$

So, while the input constraint binds, q_t converges monotonically to

$$q^* \equiv \frac{\beta\theta}{\lambda g_R},$$

and $g_{At} \to \lambda g_R/\beta$.

If $\gamma > \beta$, the parallelizability ceiling binds only when

$$A_t^\gamma < R_t^\lambda \quad \Longleftrightarrow \quad q_t < A_t^{\beta-\gamma}.$$

A_t is increasing and $A_t^{\beta-\gamma} \to 0$, while in the ceiling-binding region

$$g_{At} = \theta A_t^{\gamma-\beta}$$

is eventually large enough that $g_{qt} > 0$. So the economy eventually enters and remains in the input-constrained region. Thus $g_{At} \to \frac{\lambda}{\beta} g_R$.

If $\gamma = \beta$, the ceiling-binding region is $q_t < 1$, and

$$g_{At} = \theta \min\left(1, 1/q_t\right).$$

If $g_R < \theta\beta/\lambda$, then $q^* \geq 1$, so the stable limiting region is input-constrained and $g_{At} \to \frac{\lambda}{\beta} g_R$. If $g_R > \theta\beta/\lambda$, then q_t eventually lies below 1, the ceiling binds, and $g_{At} \to \theta$. The equality case is immediate.

A.2 Proof of Proposition 4.1

Let

$$\begin{aligned}
x_t &\equiv [\ln(A_{it})], \\
y_t &\equiv x_t - \bar{g}_A t, \\
c_{iz} &\equiv [\Phi_z]_i \bar{g}_A \qquad \gg 0 \text{ by construction}, \\
q_t &\equiv \min_i \frac{x_{it}}{\bar{g}_{A_i}}, \\
m_t &\equiv \min_i \frac{y_{it}}{\bar{g}_{A_i}} = q_t - t.
\end{aligned}$$

For technology i, denote the z^{th} constraint B_{iz}, and observe that

$$\begin{aligned} B_{i0t} &= \theta_{i0} R_{i0}^{\lambda_i} e^{[\Phi_0]_i y_t}, \\ B_{izt} &= \theta_{iz} e^{[\Phi_z]_i y_t + c_{iz} t}, \quad z \geq 1. \end{aligned}$$

If y_t is eventually lower-bounded, all g_{A_i} are eventually equal to the input constraints: the proposition is proven. This is because otherwise, there would be some i and some sequence of times t along which A_{it} grows superexponentially; but this would not occur even in the absence of parallelizability constraints, by Davidson et al. (2026), Prop. 1.1.

If A_i were bounded above for any i, then for all z, B_{iz} would eventually be bounded above zero: a contradiction. So $q_t \to \infty$.

Choose

$$\begin{aligned} &Q \;:\; \theta_{iz} e^{c_{iz} Q} > \bar{g}_{A_i} \;\forall i, z; \qquad \underline{t} \;:\; q_t > Q \;\forall t \geq \underline{t}. \\ &L \;:\; \theta_{i0} \left(R_{i0} e^{g_{R_i} L} \right)^{\lambda_i} > \bar{g}_{A_i} \;\forall i. \end{aligned}$$

We will now show that m_t is eventually lower-bounded, and thus that y_t is.

Given t, let $i \in \text{argmin}_j \frac{y_{jt}}{\bar{g}_{A_j}}$, so that i minimizes m_t and thus also minimizes q_t. Then

$$x_{jt} \geq q_t \bar{g}_{A_j} \quad \forall j. \tag{21}$$

For all $z \geq 1$,

$$B_{izt} = \theta_{iz} e^{[\Phi_z]_i x_t} \geq \theta_{iz} e^{c_{iz} q_t} \quad \text{by (21)}$$

because $\phi_{iiz} < 0$ and $\phi_{ijz} \geq 0$ for $j \neq i$. So for $t \geq \underline{t}$ and $z \geq 1$, we have $B_{izt} > \bar{g}_{A_i}$.

Likewise,

$$\begin{aligned} B_{i0t} &\geq \theta_{i0} R_{i0}^{\lambda_i} e^{m_t [\Phi_0]_i \bar{g}_A} \\ &= \theta_{i0} \left(R_{i0} e^{-g_{R_i} m_t} \right)^{\lambda_i}. \end{aligned}$$

So for $m_t \leq -L$, $B_{i0t} > \bar{g}_{A_i}$.

So for $t \geq \underline{t}$, if $m_t \leq -L$, then $g_{A_i t} > \bar{g}_{A_i}$; so $\dot{y}_{it} = g_{A_i t} - \bar{g}_{A_i} > 0$ for each technology

i minimizing $y_{it}/\bar{g}_{A_i}$; so the right derivative of m_t is positive. It follows that for $t \geq \underline{t}$, we have $m_t \geq \min(m_{\underline{t}}, -L)$.

A.3 Proof of Proposition 5

Part 1. Choose $g \gg 0$ such that $\Phi_z g \gg 0$ $\forall z$. Let

$$a \equiv \min_{i,z} \frac{\theta_{i,z}}{g_i} \Pi_j A_{j0}^{\phi_{ijz}}, \quad b \equiv \min_{i,z} [\Phi_z]_i g,$$
$$q_t \equiv \min_i \big(\log A_{it} - \log A_{i0} \big) / g_i.$$

Observe that $\dot{q}_t \geq a e^{b q_t}$, so $\frac{d}{dt} e^{-b q_t} \leq -ab$. Since $q_0 = 1$, it follows that $e^{-b q_t} \leq 1 - abt$. Since $ab > 0$, q_t cannot be finite for $t \geq \frac{1}{ab}$. The log-technology vector field is locally Lipschitz on finite points, so finite-time noncontinuability implies that some A_i diverges. The network therefore reaches a singularity at some $T \leq \frac{1}{ab}$.

Part 2. Fix $\zeta \in \mathcal{Z}^{\mathcal{N}}$, and let $A^{(\zeta)}$ denote the corresponding technology path. The full network satisfies

$$\frac{d}{dt} \log A_{it} = \min_z \theta_{iz} \Pi_j A_{jt}^{\phi_{ijz}} \leq \theta_{i\zeta_i} \Pi_j A_{jt}^{\phi_{ij\zeta_i}},$$

while network ζ satisfies

$$\frac{d}{dt} \log A_{it}^{(\zeta)} = \theta_{i\zeta_i} \Pi_j A_{jt}^{(\zeta)\phi_{ij\zeta_i}}. \tag{22}$$

The right-hand side of (22) is locally locally Lipschitz in $\log A_j$ for all j and nondecreasing in $\log A_j$ for all $j \neq i$, so the standard quasimonotone comparison theorem gives

$$\log A_{it} \leq \log A_{it}^{(\zeta)} \quad \forall i, \quad t < \min(T, T_\zeta).$$

So $T_\zeta \leq T$.