Small Modular Infrastructure

E. Dahlgren, K.S. Lackner, C. Gocmen, G. van Ryzin

July 2012
DRO-2012-03

Small Modular Infrastructure

Eric Dahlgren Klaus S. Lackner
School of Engineering and Applied Science, Columbia University
Caner Göçmen Garrett van Ryzin
Graduate School of Business, Columbia University

July 24, 2012

Abstract

In many basic infrastructure industries – transportation, electric power generation, raw material processing, etc. – we have witnessed a trend of ever increasing unit size of technology. Several factors have driven this trend. Basic geometry – how length, area and volume scale with size – often (but not always) imply lower capital costs per unit capacity and higher conversion efficiency as one increases unit size. Reducing labor cost also drives increases in unit scale. Yet advances in automation and communication technology together with a more comprehensive cost-benefit analysis argue for a potential reversal of this "bigger is better" trend – a radical shift to a world in which efficiency of *size* is replaced by efficiency of *numbers*, in which custom built technology of massive unit scale is replaced by massive numbers of small, modular, mass-produced units deployed in parallel in single locations or distributed geographically. To make this argument, we first develop a framework to evaluate the economics of unit scale. What factors drive the decision to build big or small? And how is technology altering this choice? What kind of flexibility benefits can one obtain from small unit size? We then apply the framework to several industry sectors and argue that under plausible assumptions, many industries are nearing – or already are at – a tipping point towards radically smaller unit scale. Such a transformation would have profound implications for the structure of both established and emerging industries.

1 Why build big?

In a wide range of industries, the historical trend is toward ever increasing unit size of technology.[1] Our food, once produced on small family plots, now comes overwhelmingly from large industrial factory farms. Ships that in the early twentieth century carried 2,000 tons of cargo have been replaced by modern container ships that routinely move 150,000 tons. Coal-fired power plants that averaged 50 MW of output in 1950 today approach 1 GW. The list goes on. For example, Figure 1 depicts the evolution of total installed capacity together with average generator sizes of several different electricity generating technologies in the US[2]. The data suggest that growth in a certain generation technology is accompanied by increasing unit sizes.

What underlies the trend of "bigger is better?" Before exploring this question further, we need to distinguish between the traditional notion of economies of scale, which encompasses all possible benefits that are associated with increasing total firm-wide output, and those benefits that are directly

---

[1] By *unit size* we mean the capacity of a single unit of technology, e.g., the number of people carried by a single aircraft, the MFLOPS of a single CPU, the watts of electric power produced by a single generator, etc.

[2] The explanation for the stagnation in natural gas-derived capacity in the late 70s is the enactment of "The Powerplant[sic] and Industrial Fuel Use Act", by U.S. Congress in 1978, effectively banning new construction of natural gas-fired power plants [65]. This act was repealed in 1987. The data include all generators with a nameplate capacity greater than 1 MW, meaning that once the growth stagnates in a technology small, niche installations are still built, which explains the apparent decrease in average unit sizes in coal and hydro power.

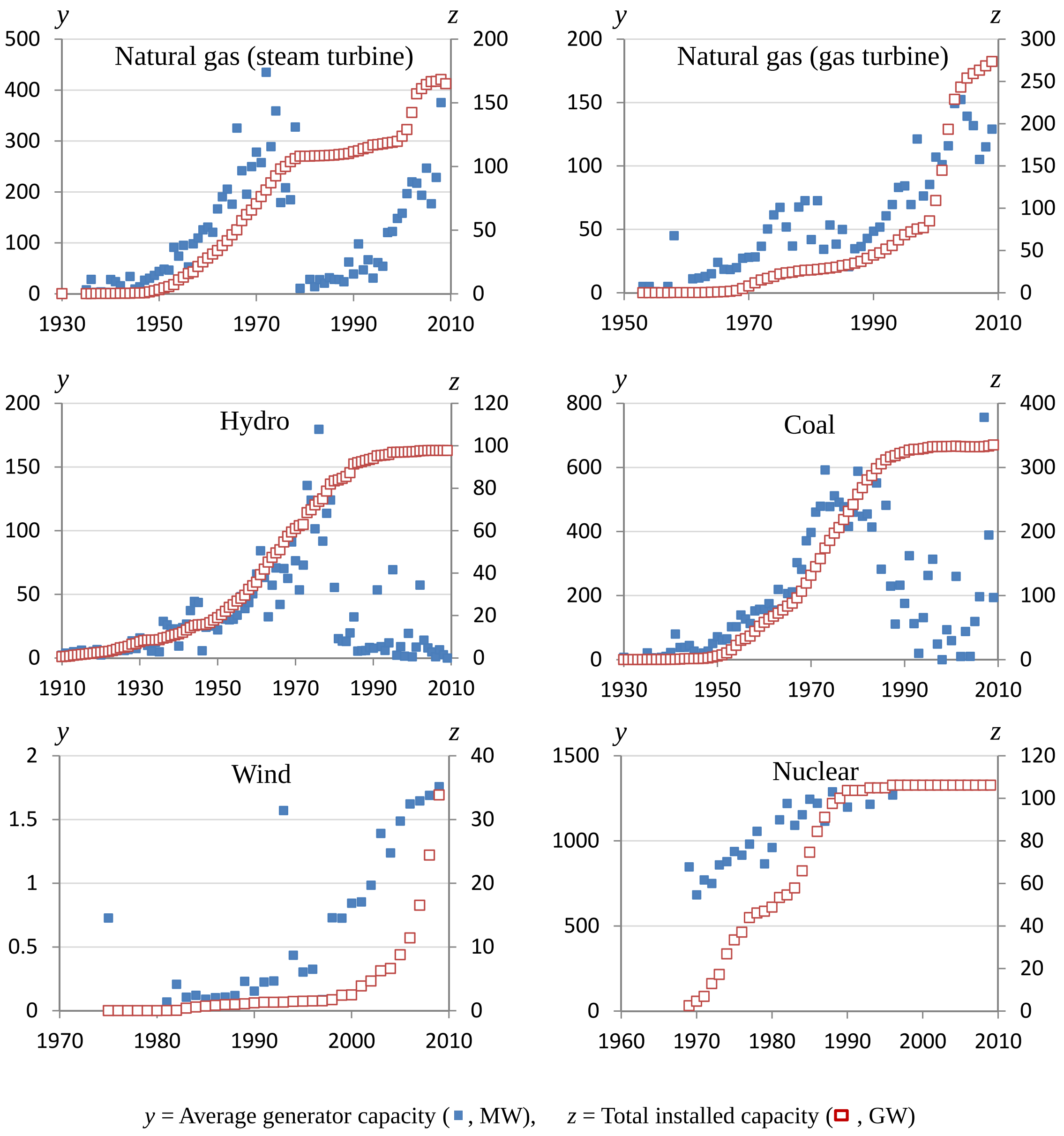

Figure 1: *The graphs show the average capacity (left axis, 'y') of electricity generators installed in the US and the total capacity (right axis, 'z') of the same technology over time [29]. Together, these technologies represent more than 90% of total generating capacity (1.1 TW) in the US in 2009. 'Natural gas (gas turbine)' includes both stand-alone gas turbines and gas turbines in a combined cycle setting, similarly for 'Natural gas (steam turbine).'*

attributed to building and operating larger individual units. Realizing that many of the benefits arising from economies of scale are actually independent of the sizes of the individual units of production, e.g. spreading out fixed overhead costs, we are here interested in those that do depend on unit size. We refer to these unit-size dependent benefits as economies of *unit* scale.

Conventional wisdom in cost engineering holds that capital costs per unit of capacity decline with increasing unit size. This empirical observation can be attributed to several different factors. Cost efficiencies arise from spreading out the fixed-cost components of a system, like control and monitoring subsystems, which must be included in a unit's design regardless of unit size. Furthermore, the costs incurred during the design phase are, to a first approximation, independent of the size of the unit. Lastly, it is commonly believed that the material required, and thereby also the cost of constructing a unit scales like the ratio of surface area to enclosed volume, which would lead to decreasing cost per unit output as size increases (see §3.1.1).

Another key driver of large unit scale is operating labor. In general, fixed quantities of labor are required to operate and maintain each unit, and hence, the total labor required to produce a given amount of output declines with the number of units used to produce that output. For example, a fleet of 100 10 tons dump trucks for hauling iron ore requires more labor (drivers) than a fleet of 10 100 tons dump trucks. Likewise (though less obviously) a power plant comprising 100 1 MW generators generally requires more labor than a plant consisting of only one 100 MW generator. Therefore, increasing unit size leads to increased labor productivity and consequently lowers operating costs per unit output.

A further motivation for increasing unit size is conversion efficiency. These efficiencies are again due in large part to geometric scaling laws. Take thermal power generation, in which chemical or nuclear energy is transformed into heat. Considering any contained part of the thermal cycle in a power plant (e.g. the boiler or a segment of the working fluid) the capacity is related to heat content, which approximately scales with volume, and dissipative heat losses scale with the surface area. In this way, larger unit size leads to lower dissipative losses and hence higher conversion efficiency.

Lastly, larger units by their very nature create centralized points of production, which can act to bring down fixed operating costs per unit produced. These include such costs as security, administration, and the transport infrastructure needed to acquire raw materials and distribute outputs, such as shipping docks, pipelines and railroad tracks.

The issue of how economies of unit scale can reduce capital and operating costs has been analyzed in the economics and engineering literature; see for example [38, 40, 52, 54, 43], and [68].

1.1 The case for small

While this development of ever larger unit size may have made sense historically, we submit that the incentives today for continuing the trend are less compelling - and indeed there may be tremendous benefit in reversing it. It is now realistic to consider a radically new approach to infrastructure design, one that replaces economies of unit scale with economies of numbers, that phases out custom-built, large-scale installations and replaces them with large numbers of mass-produced, modular, small-unit-scale technology – operated in either centralized or distributed fashion – offering new possibilities for reducing cost and improving service. In the context of electricity generation, some of these concepts are presented in the "small is profitable" work of Lovins [55], but the idea applies much more broadly. Indeed, we are already seeing such massively parallel infrastructure strategies emerging in nascent form in several industrial sectors (see §2) and, in our assessment, many other sectors are ripe for similar change (see §5).

This shift mirrors a similar revolution that began thirty years ago in the computer industry. The traditional approach to achieving high capacity and increased speed in computing was to build increasingly powerful, specialized machines with greater and greater processing power. This trend, however, came to a dramatic halt in the mid-1990s. At this point, it became cheaper to employ mass-produced CPUs and high capacity memory from the burgeoning personal computer and workstation industry. Large numbers of small-scale CPUs could achieve high computing capacity without relying on ever-more-powerful single-processor machines [71].

Like supercomputers in the 1990s, we believe that many basic infrastructure industries are near or at a tipping point from large to small optimal unit scale. And three driving forces underlie this technological shift.

First, technologies for automating processes exist today that were previously unavailable. In the past, a massively parallel modular plant was simply infeasible because of excessive personnel cost. In contrast, current computing, sensor and communication technologies make high degrees of automation possible at very low cost. The result is a radical undercutting of the logic that significant labor savings can only be obtained through large unit scale.

Second, as we argue below, mass production of many small standardized units can achieve capital cost saving comparable to, or even larger than those achievable through large unit scale. For instance, a mass produced car engine costs \$10/kW, while a typical large-scale coal-fired power plant costs about \$1000/kW [51]. While this possibility is hardly "new," by chasing unit scale so relentlessly, many infrastructure industries have simply missed out on the cost-reducing potential of mass producing capital. Moreover, since operating labor cost alone rendered small unit scale technology uneconomical in the past, there was little incentive to pursue the possibility of mass-produced capital. Today, that situation is fundamentally altered.

Lastly, there are many inherent flexibility benefits to small unit scale technology which, in the past, have largely been ignored in the race toward ever increasing scale. Small-scale units can be used in multiples to better match the output requirements of a given project and can also be deployed gradually over time, both of which reduce investment cost and risk. They also offer geographic flexibility; multiple small units can be aggregated at a single location to achieve economies of centralization (e.g. to reduce overhead or transport costs) or they can be distributed to be closer to either sources of supply or points of demand. An additional benefit of small unit scale is flexibility in terms of operating output; having many units of small scale makes it possible to selectively operate varying numbers of units to better match short-run variations in demand. Also, with small unit scale technology, one can achieve high reliability through enormous redundancy and statistical economies of numbers.

1.2 Overview

To further explore the possibilities of small modular infrastructure, one must look at existing small-scale modular technologies, the determinants of economies of scale, the impact of learning curves on capital costs and the effect of unit size on operating costs. It is also important to account for the many flexibility and diversification advantages of small modular units. Toward this end, in §2 we look at existing examples of small modular technology. §3 then provides theory on how economies of unit scale and learning affect capital costs, and there we discuss the effect of unit size on operating costs. §4 looks at the flexibility advantages that come with employing small modular units, such as investment flexibility and diversification. Afterwards, in §5 we give examples of several technologies where the trend of ever increasing size has been observed but which have the potential to benefit from smaller unit scale. Lastly, §6 concludes with some general observations about how we can make the transition to "thinking small".

2 Existing industry examples

Before analyzing the case for small unit scale in more depth, it is helpful to consider some current industry examples. Small, modular nuclear reactors, chlorine plants, and biomass gasification systems are technologies that are either already commercially available or currently under development, and importantly, these technologies have taken advantage of small unit scale and the economies of mass manufacturing. The specific technologies are very different, and they face different impediments to a large unit scale; regulatory hurdles for nuclear power, safety hazards for the production and transportation of chlorine and the distributed nature of the inputs for biomass gasification. In each case, these reasons were enough to tip the scales in favor of a smaller unit scale. However, regardless of the technology, the execution of the new strategy is similar: mass produce and automate.

2.1 Small modular reactors (SMRs)

Small-scale applications of nuclear power may seem counter-intuitive, but there is now growing interest in a new generation of small, modular nuclear reactors (SMRs). Unlike existing generators which have an average capacity close to 1 GWe, SMRs have a capacity as small as 25 MWe [8, 18]. Because of their small unit scale, reactors can be manufactured off-site and transported to the power plant location semi- or fully-constructed; some are even small enough to be transported by truck. As a result, the time required to bring a new plant on line – one of the biggest obstacles inhibiting the adoption of nuclear power – is shortened from an average of twelve to less than five years [46, 26]. Along with long lead time, high capital costs have been a major impediment to building nuclear power plants, with a typical large-scale nuclear power plant requiring a minimum investment of around $12 billion in capital [76]. With SMRs, the upfront capital cost can be as low as $100 million. Moreover, investors can easily expand a plant in the future depending on market conditions [12]. Scalability of SMRs also makes nuclear power attractive for smaller projects, thereby increasing the market for such technology. For example, SMRs can be used as "drop-in replacements" in aging power plants since they can utilize the existing transmission capacity or in remote applications like mining and oil and gas production where dependable baseload power is crucial [12, 2].

While SMRs can be thought of as scaled down versions of classical reactors, they share a common set of design characteristics which make them safer than classical reactors [49]. Consequently, safety and support systems are less complex. In addition, due to their modularity, major maintenance tasks can be handled by shipping the entire reactor unit back to the factory. These factors can reduce the on-site staff size by a factor of five, per unit of power output, compared to traditional nuclear power plants [46]. Another operational advantage of SMRs is the reduced impact of maintenance down-time. Current nuclear power plants typically have one or two reactors, so during maintenance they lose most, if not all, of their generating capacity. With SMRs, a typical power plant would have 4-12 reactors, so if one reactor is taken off-line for maintenance, the impact on total generation capacity is much less [46].

There are numerous small modular reactors being designed around the world. Examples include Toshiba's 4S, Babcok & Wilcox's mPower, Hyperion's HPM, and Rosatom's KLT-40. Currently, most of the designs are being approved, and the first SMR plant is expected to be operational by 2018 [67]. There are also plans to use SMRs in unconventional settings. For example, Rosatom, a Russian state corporation in charge of the nuclear complex, has ordered the construction of several floating nuclear power stations. The first, Akademik Lomosonov, was launched in 2010 and is expected to become operational by 2012 [70]. Another example is the small offshore reactor Flexblue developed by DCNS, a French Naval shipyard company. The first prototype is scheduled for 2013 and is planned to enter commercial production by 2016 [14].

Nascent SMRs provide a sense of how a radically smaller unit scale can fundamentally disrupt an industry accustomed to a massive scale. Their designs are simplified and standardized; they are manufactured in the factory and not in the field; investment is significantly more flexible and less risky; and their small size opens up entirely new domains of applications for nuclear power. All these are key advantages of the small modular approach to technology.

2.2 Small modular chlorine plants

Chlorine is a chemical element used widely for the production of industrial and consumer products, disinfection and water purification. Being highly toxic, chlorine is dangerous and costly to store and transport. For example, in a highly publicized train crash in 2005 in Graniteville, South Carolina, rail tank cars carrying chlorine ruptured resulting in the death of nine people and the evacuation of thousands of residents [24]. To avoid the costs and risks associated with chlorine storage and transportation, companies such as GE, MIOX and AkzoNobel have come up with designs for small skid-mounted modular chlorine plants which can be placed close to sources of demand. Manufacturing these modular plants off-site reduces on-site construction requirements, shortening the time needed to bring a plant online. Modular design also enables efficient inspection and maintenance, reducing operational costs. All designs are highly automated, which decreases the requirement for on-site skilled personnel. Moreover, AkzoNobel's design aims to completely eliminate on-site personnel requirement by controlling the facilities remotely, thus realizing the scale benefits of traditional chlorine plants by providing centralized monitoring and control.

The idea for small modular chlorine plants has been around for some time. GE's Cloromat has been in commercial use for over 35 years and, MIOX has been producing its own line of plants since 1994 [3, 5]. AkzoNobel's solution is quite new, with the first plant targeted to come online in 2012 [7, 17].

Again, this trend of small modular chlorine plants hints at what is possible more broadly with a strategy of small unit scale. Production is distributed and located close to points of demand, eliminating the need for transportation. Plant designs are standardized and mass produced to reduce capital costs. And on-site labor is minimized - or even eliminated - by utilizing advanced sensing, automation and communications technology that enable remote control of plant operations. The overall approach achieves many of the capital and labor savings of large unit scale yet provides benefits, such as distributed operation, that can only be achieved with small plants.

2.3 Small modular biomass gasification systems

Biomass, e.g. wood and agricultural waste, is a widely available renewable energy source that can be used to produce electricity and heat through gasification. Being abundant, biomass is a good candidate for use in distributed electricity production and accounts for approximately 1-2% of U.S. electricity production [11].

Several companies such as AESI, Community Power Corporation(CPC) and Innovative Energy Inc.(IEI) have commercially available designs for small modular gasification devices. These devices convert biomass into synthesis gas to generate electricity and heat. AESI's and CPC's units have capacities in the range of 50-100 kWe and IEI produces standard 1-2 MWe units [9, 6, 4]. As a consequence of their small size, AESI and CPC ship their systems housed in a small number of shipping containers which are connected to each other on-site. This modular approach allows for easy integration and shortens the time required to get a system operational. IEI's system is larger, and it is also brought to the site semi-manufactured where it is connected to existing units and infrastructure. All designs are highly automated and CPC's design allows its equipment to be controlled remotely from a central location over the internet.

Like modular chlorine plants, small modular biomass gasification systems have already been deployed in numerous projects. IEI is constructing a 5 MWe waste-to-energy power plant in Missouri, and AESI has deployed its system at a pharmaceutical plant in North Carolina [27, 72]. CPC lists numerous institutions, such as Kedco, Shell Solar, Idaho Power Corporation and Western Regional Biopower Energy Program, as its customers [1].

Again, this example of small modular biomass plants illustrates a broader strategy of combining mass manufacturing and highly automated operations to enable small unit scale plants to achieve economies of capital and operating costs comparable to their larger brethren, while creating entirely new benefits such as short lead times and distributed operation only achievable by small scale technology.

2.4 Other Examples

Other existing examples of small modular technologies include, supercomputers, data centers, and small steel mills. One of the most prominent examples of supercomputers that use small modular units is the Condor Cluster, which is owned by the U.S. Air Force and makes use of off-the-shelf Playstation 3 gaming consoles. By employing this mass produced computing equipment the Air Force was able to reduce capital costs more than 20 times [81].

Data centers are another important setting where we can observe small modular technology. With the widespread use of Internet, the need to house data in centralized locations became important. Data centers achieve this task by using standardized modular equipment in parallel. This allows them to efficiently scale up or down depending on market conditions.

Lastly, this trend can be seen in steel production. A traditional integrated steel mill makes steel from iron ore using a blast furnace. By using much smaller steel mills, which employ electric furnaces, it is possible to produce steel from scrap metal at a much lower cost. These so called mini-mills are more efficient in processing scrap metal then conventional large steel mills, and are also less costly to start and stop which enables them to manufacture steel in small batches.

3 A theory of unit scale

We next lay out a theory that explains the fundamental factors that drive the choice of unit size, and how they are changing due to advances in technology and the shifting demands of the global marketplace. We argue that there can be significant advantages to radical changes in unit scale, but in order to do so, one must fully understand the motivation for large-scale units.

3.1 Capital costs

Given a prototype unit of known cost, one can follow two different strategies in order to produce output that greatly exceeds that of the prototype. One option is to create a single large unit of sufficient capacity; in effect, simply scaling up the size of the prototype. The other option is to provide a large number of standardized units that aggregate to produce the total desired output. Such a massively parallel production strategy is possible as long as there is no significant cost to combining the separate outputs into a single stream.

Capital cost, regardless of strategy, is an important determinant of the economic viability of any project. The literature on cost engineering offers empirical relations between the cost of the required equipment and the production capacity of the system for both strategies. In this section, we compare these costs and conclude that, based on established empirical relationships, the cost per unit of capacity approximately scales the same in both cases. Hence, we argue that capital cost considerations are not likely to determine the optimal size of the production equipment. In the following, we will refer to the cost reductions achievable in the two cases as the economies of unit scale and economies of mass production, respectively.

3.1.1 Economies of unit scale

A traditional method of estimating the cost k of a piece of equipment with capacity c uses a power law:

$$k(c) = k(c_0)\left(\frac{c}{c_0}\right)^{\alpha}, \tag{1}$$

where $k(c_0)$ is the cost of a reference unit of capacity c_0. If the exponent is less than unity ($\alpha < 1$), the cost per unit size is decreasing, ($d(k(c)/c)/dc < 0$), creating an impetus for building larger units. Numerical values for α have been estimated for a wide array of process equipment. Typically these range from 0.6 to 0.8 [40, 44, 32], hence the so-called "0.7 rule", or sometimes "two-thirds rule." It is worth noting here that there are practical limitations to how large one can build working structures, since the structural integrity of materials becomes a factor for very large sizes. Consequently, pushing the boundaries on the large end of the spectrum has normally been accompanied by development in materials that are lighter and stronger, as for example in the use of carbon fibers in wind turbine construction. (See [52].)

The observed decline in cost with increasing unit size of a piece of equipment can be attributed to several factors mentioned above. However, a frequently cited argument for the appearance of the scaling law in (1), with values of the exponent α ranging from 0.6 to 0.8, is based on the geometric relationship between surface area and volume [37, 41, 74, 59]. This argument suggests that costs scale with the amount of material used in the structure, which in turn is supposed to scale with the surface area. Often when considering a piece of industrial equipment like a pressure vessel, a chemical reactor or a truck bed, it is reasonable to assume that its capacity is proportional to its useful volume. If both assumptions apply, scaling the capacity with a factor λ results in costs scaling as $\lambda^{2/3}$. This offers a nice explanation for the often-observed value of $\alpha \approx 2/3$ in (1).

However, from the perspective of structural mechanics, this argument is flawed. In most situations, as the volume of a structure is increased, it is necessary to increase the wall thickness as well in order to preserve structural integrity. As a result, the mass of an optimally designed unit typically increases far more rapidly than the 2/3 power of the enclosed volume. Indeed, in any situation in which the weight of the structure matters, it is usually not even possible to achieve symmetric scaling, i.e., a scaling of the structure in which all linear dimensions increase by the same scaling factor, $\lambda^{1/3}$. Instead, wall thicknesses or diameters of structural members grow faster than the linear size of the system. Therefore, the mass of the unit, or the amount of material in the walls of the unit, and thereby also costs, would scale faster than the capacity of the system.

In living systems, the break-down of symmetric scaling can be found by comparing a mouse to an elephant, where the latter has to have disproportionately thicker legs to support its weight; the same concept holds for industrial objects. A structure operating at its mechanical capacity (with appropriate safety factors) cannot be uniformly scaled up. The weight of the larger structure would exceed the limits of its structural integrity. The way around this problem is to use lighter materials, stronger materials or a combination of the two. Typically, these materials are not used at smaller scales because they are more expensive. However, even with the most advanced materials, physics imposes a boundary which can only be pushed so far. Appendix §A provides a detailed mathematical analysis of the scaling of solid structures.

While we do not challenge the observed empirical relationship between cost and unit size, we submit that the conventional explanation, which suggests that the wall area to volume ratio drives scaling behavior, is overly simplistic and under more careful analysis proves incorrect. Quite the contrary, we argue in the appendix that structural relationships tend to reduce cost advantages of larger units, and, if they were to dominate, the overall cost structure would in fact result in diseconomies of unit scale.

3.1.2 Economies of mass production

The economies of mass production was first investigated by Wright [80] in the context of airplane production. The seminal work by Arrow [20] provided the first general analysis of the subject in which he argues that costs of manufactured goods decline with the cumulative number produced. There have since been various studies resulting in ample data on the cost reduction as the cumulative production grows (see e.g. [19, 33, 58, 75]).

It is no surprise that costs decline as cumulative output increases. First, efficiencies improve dramatically due to both specialization and improved product design. For example, when organizing a production process to manufacture a large number of identical units, one can justify high degrees of specialization in tools, layout and job design that can dramatically reduce per-unit manufacturing costs. Also, in high-volume manufacturing, it becomes justifiable to invest significantly in product design in order to make parts and subsystems more integrated, easier to assemble and hence less costly to make (*design for manufacturing*).

The second benefit of producing in volume is learning. As a manufacturer gains cumulative experience producing a given product via a certain process, myriad improvements in design, materials and production methods are uncovered. Such a process of *continuous improvement* can lead to significant cost reductions as production volumes increase. Conversely, cost reductions attributed to learning can reverse themselves given extended breaks in production. This trend, akin to a "forgetting curve", can explain the relatively small cost reductions over time for larger and more long-lived installations. Importantly, installations endowed with greater longevity also tend to be custom-made rather than mass-produced which further contributes to the comparably small reductions in cost from one investment to the next [58].

The effect of declining cost with the number of units produced is commonly formulated using learning curves, which state that the unit cost decreases by a fraction $\varepsilon < 1$ as the cumulative production doubles. That is, the cost, k_{2n}, of the $2n$-th unit is a fraction $\varepsilon < 1$ of the cost, k_n, of the n-th unit, or

$$\frac{k_{2n}}{k_n} = \varepsilon. \tag{2}$$

Sometimes this cost reduction is expressed by the learning rate, defined by $1 - \varepsilon$. Based on (2), a continuous approximation of k_n can be formulated as

$$k_n = k_1 \varepsilon^{\log_2 n} = k_1 n^{\log_2 \varepsilon},$$

where k_1 is the cost of the first unit produced. The aggregated cost, $K(N)$, of N mass-produced units following the given learning curve, can then be expressed as

$$K(N) = \int_1^N k_n dn = \frac{k_1}{1 + \log_2 \varepsilon} N^{\log_2 \varepsilon + 1}. \tag{3}$$

We note that even though this reduction in cost is referred to generally as a "learning rate", we do not wish to imply that learning is the only contributor to this form of cost reduction; the aforementioned economies of specialization also play a key role.

3.1.3 Comparing economies of unit scale with economies of mass production

The expressions in (1) and (3) offer cost estimates of the distinctly different strategies of producing systems of large total capacity. To compare the two, we consider the options of either scaling up a reference unit of capacity c_0 and cost k_0 by a factor N (economies of unit scale) or manufacturing N copies of the same reference unit (economies of mass production). Either way, the end result is a system of total capacity Nc_0. The two cost estimates yield

$$\begin{aligned} k(Nc_0) &= k(c_0)\left(\frac{Nc_0}{c_0}\right)^{\alpha} = k(c_0)N^{\alpha}, \qquad \text{(Economies of unit scale)}, && (4) \\ K(N) &= \frac{k_1}{1+\log_2 \varepsilon} N^{\log_2 \varepsilon + 1}, \qquad \text{(Economies of mass production)}. && (5) \end{aligned}$$

A statistical analysis performed in [33], based on a sample of 22 different mass-production-oriented industrial sectors found an average learning rate of 19% ($1 - \varepsilon = 0.19$), which corresponds to a value of the exponent in (5) of 0.7, i.e. $\log_2 \varepsilon + 1 = 0.7$. Since typical values for the exponent α in (4) range between 0.6 to 0.8 it is reasonable to conclude that $\log_2 \varepsilon + 1 \approx \alpha$ and hence, the reductions in production costs ensuing from economies of mass production are on par with those from economies of unit scale.

It should be noted that this comparison assumes that the total installed capacity is independent of the choice of unit size. This would be true, for example, in building a single factory. However, smaller unit sizes can open up new domains of application for a given technology, and hence increase the overall market size for that technology. This, in turn, can further reduce cost per unit capacity. For example, as noted in §2.1 current sizes of nuclear reactors make them infeasible for a wide range of applications. With the introduction of SMRs, nuclear energy may become a viable option for much smaller projects.

3.2 Operating costs

As noted previously, high conversion and labor efficiencies are traditional benefits of large unit scale production. However, the degrees in which these factors affect cost are changing dramatically with recent advances in technology. Given these developments, one must take a closer look at the benefits of scale and how technology can capture these benefits without resorting to large scales.

3.2.1 Labor efficiency

The labor required to operate a given process can roughly be divided into variable and fixed labor. Variable labor scales with total output whereas fixed labor scales with the number of individual units employed in the process, e.g., machine operators and maintenance personnel. For purely illustrative purposes, the labor force on a commercial airliner, slightly simplified as flight attendants (variable labor) and pilots (fixed labor), can serve to clarify the distinction. Within reasonable limits, the number of pilots required on an aircraft is independent of the unit scale (number of passengers that can be carried), whereas the number of required flight attendants depends on the number of passengers carried. The distinction in types of labor is usually less clear-cut than in this simple example, but the definition is nonetheless useful.

Processes where fixed labor cost is a significant component of total cost indicate the potential for increasing labor productivity by scaling up individual unit sizes. Arguably, this has been a common trait in many energy and materials processing industries in the past century, thereby driving, at least in part, the trend of ever increasing unit size. A notable example can be found in the mining industry and its pursuit of ever-larger unit sizes in extracting equipment, as detailed in §5.3. However, such a strategy for increasing labor productivity will eventually run into physical barriers that are progressively harder to surmount.

An alternative strategy to reduce labor cost is to employ automation which can ultimately drive labor cost to zero, or at least decouple it from the number of individual units employed. Naturally, every process has its own characteristics and its amenability to automation needs to be evaluated on a case-by-case basis. However, the progress made in wireless communication, GPS technology, sensor technologies and computational processing power is fundamentally changing the economics of automation; the capabilities of these technologies are soaring while costs are plummeting, enabling unprecedented degrees of low-cost automation.

To be specific, Nordhaus [60] reports that the average price of computing power decreased by more than 40% per year between the years 1990-2002. The communication sector has also seen significant advances. International Telecommunication Union (ICT) reports that the average price for a high-speed Internet connection dropped by 52% in the world between the years 2008-2010, and Akamai reports that average global peak connection speed increased by 67% between 2010-2011 [15, 47]. The result is that automation and remote sensing and control technologies now provide tremendous capability at very low cost.

Several well-known instances where such automation has flourished are electronic toll collection, automated warehouses and electronic check-ins for flights. In several instances, such as with ATMs and car-sharing services such as Zipcar, automation has changed industry dynamics by making it possible to serve areas and markets which were previously too costly. While automation does not eliminate all the benefits of increasing unit scale, given the current state of technology, automation is often much cheaper than employing human workers, and this reduces the benefits of large unit scale significantly.

Two other strategies to decouple the labor cost from unit scale and the geographical proximity of units are remote operation and centralized maintenance. With appropriate instrumentation and automation technologies linked to the Internet, a central control center can monitor and operate units remotely, eliminating the need to have on-site personnel at every location and increasing utilization of operators due to pooling economies. As a result, labor costs become independent of the physical proximity of the units, and there is less incentive to keep units together geographically. Similarly, small units requiring maintenance can be shipped back to the manufacturer or to a local maintenance center for major repairs and upgrades. Again, this eliminates the need for a local maintenance staff and also creates pooling economies in maintenance and repair labor. The examples of SMRs and modular chlorine plants described in §2 illustrate these ideas in practice. When combined with automation technology, remote operation and centralized maintenance enable small unit scale technology to achieve the levels of fixed and variable labor cost previously obtainable only by centralization and massive unit scale.

3.2.2 Conversion efficiency

Conversion efficiency is the ratio of inputs to outputs, for example how much energy or raw material is required to produce a unit of output. The same geometrical ratios that favor larger unit scales in terms of material costs often imply high conversion efficiency too, such as the reduction of thermal losses from working fluids discussed previously. In a similar way, geometry also influences the efficiency in turbines and compressors where the main frictional losses occur along the spinning structure's circumference and the circumference scales linearly with the size of the turbine while the power output scales proportionally to the spinning structure's area, which grows like the square of the size. Hence, the output power grows faster with increasing size than the majority of frictional losses. Other scaling efficiencies arise from wasted materials in batch production processes of compounds (specialty chemicals, pharmaceuticals, cosmetics, etc.), since residue waste tends to grow like the area of a vessel, while the output capacity grows like the vessel's volume. So again, larger unit scale tends to improve conversion efficiency of these processes.

While one cannot disregard these geometrical arguments for conversion efficiency due to large unit scale, these arguments provide only a guideline and need to be evaluated on a case-by-case basis. For example, thermal power generation, previously discussed as a case for increasing unit size, does indeed seem to favor a larger size when considering constant operation. However, it is less clear how significant these size-dependent conversion efficiencies are in practical operation, e.g. a car engine operated under optimal conditions can exhibit conversion efficiencies in the range 30-

35% which is similar to those of large single-cycle power plants [78]. Another such example is water desalination using reverse osmosis explained in §5.2. Small-scale desalination systems intended to provide fresh water onboard recreational sailboats exhibit specific power consumption on par with modern utility-scale plants. Moreover, conversion efficiencies are primarily important when inputs are costly, such as fossil fuels or other purchased feed stocks. When inputs are sourced from the ambient environment, and hence effectively "free", as is the case for example with renewable energy technologies, the importance of efficiencies is greatly reduced.

In summary, conversion efficiency clearly plays a role in deciding between large and small scale unit production. However, one must take into account the variation exhibited under suboptimal operating conditions and the cost of the input sources rather than relying simply on accepted doctrine.

4 Flexibility and diversification

In addition to merely countering the arguments that traditionally have supported ever increasing physical size of capital equipment based on capital and operating costs, there are inherent flexibility and diversification benefits that can be attained only at a smaller unit scale. A careful examination of these benefits shows that they can be highly significant and easily tip the scale in favor of small unit scale.

4.1 Locational flexibility

Unlike large unit scale technologies, small unit scale offers the option of either centralization or decentralization. Multiple units can be aggregated at a single location to achieve economies of centralization in, for example, pooling the risk of demand variation. Alternatively, units can be distributed closer to either sources of supply or points of demand and thereby reduce transport or transmission costs of either in- or out-bound goods. This trade-off is different when the source of inputs are themselves localized (e.g. coal mines) versus ambient (e.g. wind and sunlight) or when demand is centralized versus distributed.

The following simple model illustrates how decentralization is affected by the scale of the production unit. Consider the case of a firm that produces a good with a demand uniformly distributed throughout a large service area. This means that the size of a production facility is proportional to the area that it serves and hence, the capital cost of the plant can be seen as a function of this area. Another area-dependent cost is the transportation cost of the output from the plant to the point of consumption. In this example, all other costs per unit output produced are assumed to have little or no dependence on the area served by a single facility, and therefore, they need not be considered in an optimization of the service area of a single plant. The question facing the firm is how large the individual production facility should be, or equivalently, how large of an area, A, each facility should serve. In this context, the firm's goal is to minimize the area-dependent cost per unit output, $K(A)$, where

$$K(A) = K_T(A) + K_C(A), \tag{6}$$

and where $K_C(A)$ and $K_T(A)$ represent the costs per unit delivered that can be attributed to capital expenditure and transportation respectively.

The transport cost per unit of output, $K_T(A)$, is a function of the shipping distance and the rate of transport. In transporting outputs over larger distances a firm is able to take advantage of economies of scale and thus, the cost of transport per unit of output will typically grow less than linearly in the shipping distance. On the other hand, it is usually reasonable to assume that the cost of shipping a unit does not decrease with increasing shipping distance. Assuming that $K_T(A)$ follows a power law,

$$K_T(A) = K_T(A_0)\left(\frac{A}{A_0}\right)^{\beta}, \tag{7}$$

then these arguments about scaling with distance imply $0 < \beta \leq \frac{1}{2}$. (See, e.g. [50].) The multiplier $K_T(A_0)$ represents the unit shipping cost for a service area A_0.

When considering only one individual unit we can use the capital cost estimate stemming from economies of unit scale, discussed in section 3.1.1. Denoting the uniform demand density by D, the size of a unit serving an area, A, is then DA, and from (1) we estimate the cost of the plant, $k(DA)$, using a power law,

$$k(DA) = k(DA_0)\left(\frac{DA}{DA_0}\right)^{\alpha}.$$

We can relate this expression of the capital cost of the plant to the capital cost per unit output produced, K_C, through

$$K_C(A) = \eta\frac{k(DA)}{DA} = K_C(A_0)\left(\frac{A}{A_0}\right)^{\alpha-1}, \tag{8}$$

where the proportionality constant, η, contains the conversion from the capital cost of the plant to the contribution to a single unit of output. This proportionality constant involves many parameters including the lifetime of the plant. For typical values of α (between 0.6 and 0.8), the function $K_C(A)$ in (8) is decreasing in the area, A, which is to be expected from the economies of unit scale. Combined with the transportation cost per unit of output, $K_T(A)$, which typically is an increasing function of the area served, the total area-dependent cost, $K(A)$, in (6) is convex with an optimal area giving the minimum cost. Furthermore, for a small enough area the cost $K(A)$ is dominated by capital cost and conversely, for large areas $K(A)$ is dominated by transportation cost. As shown in appendix §B, analyzing the assumed convex function, $K(A)$, reveals some non-trivial conclusions. For example, the simple heuristic of sizing the plant such that the capital cost equals the transportation cost will give a total area-dependent cost at worst a factor 2 greater than at the optimum area.

Instead of considering the costs associated with a single plant we can alternatively consider the costs associated with servicing a given area, A_{total}. Now we can include the possibility of having many small plants, each servicing an area A, which together serve A_{total}. That is, we have $N = A_{\text{total}}/A$ plants. The economies of mass production described in section 3.1.2 suggest that the capital cost decrease with the cumulative number of plants built, N. The capital cost per unit output delivered in A_{total} therefore differs slightly from (8) which considers only one plant. Instead we now have

$$K_C(A) = \eta\frac{k(DA)}{DA_{\text{total}}}N^{1+\log_2\varepsilon} = K_C(A_0)\left(\frac{A}{A_0}\right)^{\alpha-(1+\log_2\varepsilon)}, \tag{9}$$

where the exponent on $N = A_{\text{total}}/A$ comes from (3). As discussed in section 3.1.3, observations suggest that the exponent in (9) is close to zero. If the economies of mass production are sufficiently strong or if economies of unit scale are weak, so that $\alpha-(1+\log_2\varepsilon)\geq 0$, then $K_C(A)$ is an increasing function of A and since the transportation cost is assumed to be increasing, the total area-dependent cost, $K(A)$, is now increasing as well. In this situation, the optimal area is $A = 0$, and the system is driven to the extreme of small and distributed units.

An example that demonstrates the benefits of decentralization is distributed electricity generation as noted by Lovins [55]. According to [10], on-site generation could result in 30% cost savings in transmission and distribution, which together account for above 40% of the cost of electricity for residential customers [63]. Furthermore, distributed generation allows for the combined generation of heat and electricity which can result in energy savings from 10% to 30%, depending on the type and size of co-generation units [63].

Decentralization also has safety and security implications. For example, while small nuclear reactors (SMRs) may increase the domain of applications for nuclear energy, dispersed nuclear generation offers a greater number of targets for individuals with disruptive motives. Another concern is nuclear proliferation. With widespread adoption of nuclear technology, it becomes more difficult to monitor nuclear fuel to ensure proper handling and secure distribution. Yet these security risks can be managed by other means. For example, most SMRs are designed for off-site refueling, which reduces the accessibility of the core by unauthorized personnel. In other cases, decentralization can have the potential of improving safety and increasing security. Chlorine production is a clear example where both storage and transportation carries significant safety risks due to the toxicity of the product. Decentralization enables the production of chlorine close to points of demand thus reducing the need to store and transport this hazardous substance. This results in a substantial decrease in the safety risks associated with the use of chlorine. More generally, by its sheer nature, a smaller scale of any technology reduces the local impact of possible catastrophic failure.

4.2 Investment flexibility

Unit scale affects the flexibility of an investment in several important ways. First, small-scale units can be used in multiples to better match the output requirements of a given project thus avoiding either capacity shortages or excess capital investment. Second, small units can be deployed more flexibly over time. They can be installed sequentially as uncertain demand evolves, avoiding excess investment in the early life of a project or investment errors later in the project life cycle. Also, mass-produced, standardized units can be built to stock and deployed more quickly than custom-built, large-scale units, reducing the financing lag between investment and revenue generation. Finally, while the lifetimes of custom-built infrastructural investments tend to be very long, this need not be true for small-scale technologies. A shorter investment cycle for small-scale technologies would allow for disengagement if market conditions worsen without forsaking large sunk costs.

A full-fledged practical evaluation of these benefits would require detailed stochastic models of the underlying variables, and a rigorous real option valuation. We instead consider simple models that illustrate the significance of these investment advantages individually. The first two models are adapted from [39].

4.2.1 Investment advantages of modularity

Being able to make investments in small increments over time provides significant economic advantages. The concept is best illustrated by an example: consider a firm, such as an electric utility, planning the future expansion of a plant to satisfy increasing demand. The firm has to satisfy all demand and has two options for investment: modular and non-modular. The modular investment involves increasing the capacity in increments of x and the non-modular investment involves a one-shot investment of nx units, where $n > 0$. For simplicity we assume that n is an integer.

The current capacity matches the demand, and demand each year either increases by x with probability p or stays the same with probability $1 - p$. The lead-time for both types of investments is zero and the total additional capacity to be installed is nx. With the first increase in demand, the firm can either choose to make a one-time capacity expansion of nx units or increase the capacity in increments of x every time the demand increases, for a total of n times.

The cost of a big (non-modular) and a small (modular) investment for each increment is denoted by K_{big} and K_{small}, respectively. With a constant discount rate, r, the expected discounted cost of the non-modular investment (occurring the first time demand increases from current level) is

$$I_{\text{big}} = \sum_{k=0}^{\infty}(1-p)^k p \frac{K_{\text{big}}}{(1+r)^k} = K_{\text{big}} \frac{p(1+r)}{r+p}.$$

Turning now to the modular investment scenario, denote by I_i the expected discounted cost of the i-th investment. This investment occurs at the end of the k-th year ($k > i - 1$) when demand rises for the i-th time. That is,

$$I_i = \sum_{k=i-1}^{\infty} \binom{k}{i-1} (1-p)^{k-i+1} p^i \frac{K_{\text{small}}}{(1+r)^k} = K_{\text{small}}(1+r)\left(\frac{p}{p+r}\right)^i.$$

(For derivation of the last step see appendix §C.) Given the expected cost of each increment, we can calculate the total expected discounted cost, $I_{\text{small}} = \sum_{i=1}^{n} I_i$, of the modular scenario:

$$I_{\text{small}} = K_{\text{small}} \frac{p(1+r)}{r} \left(1 - \left(\frac{p}{p+r}\right)^n\right).$$

Taking the ratio of the total costs of the modular and the non-modular strategies we have

$$\frac{I_{\text{small}}}{I_{\text{big}}} = \frac{K_{\text{small}}}{K_{\text{big}}} \frac{(p+r)}{r} \left(1 - \left(\frac{p}{p+r}\right)^n\right) = \frac{K_{\text{small}}}{K_{\text{big}}} \left(1 + \frac{1}{\rho}\right) \left(1 - \left(\frac{1}{1+\rho}\right)^n\right), \tag{10}$$

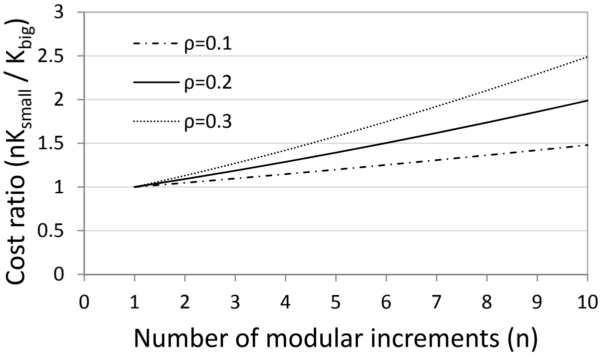

Figure 2: *The ratio* nK_{small}/K_{big} *evaluated at different* n *with equal total cost* $I_{small} = I_{big}$ *and with* $\rho = r/p = \{0.1, 0.2, 0.3\}$.

where $\rho = r/p$. The ratio $I_{\text{small}}/I_{\text{big}}$ is strictly decreasing in ρ, corresponding to increasing attractiveness of the modular investment strategy as ρ increases. The reason for this is straightforward; increasing the discount rate, r, lowers the present value of future costs and decreasing the probability of demand increase p has the consequence of further deferring these costs into the future. Both reduce the present-value cost of the modular strategy.

Demanding that the two scenarios have the same total cost, $I_{\text{small}} = I_{\text{big}}$ and rearranging (10) reveals how much more one is willing to pay for capacity nx spread out over n separate investments, i.e. nK_{small} rather than incurring all the cost, K_{big}, at once. As can be seen in Figure 2, the ratio nK/K_{big} increases almost linearly in n. For instance, with $\rho = 0.2$ as in Figure 2, considering 10 years of demand increase; the total investment cost $10K_{\text{small}}$ of the modular strategy is allowed to be over twice that of the one-off investment cost K_{big}; yet still produce an equivalent present-value of the total cost.

4.2.2 Investment advantages of shorter lead-time

Mass-produced modular technology that is manufactured to stock can significantly reduce the lead-time to deploy a new investment. We next examine how shorter lead-times can be beneficial in terms of total investment costs, again via a simple example. Similar to the previous section, assume a firm is planning the future expansion of a plant that has to satisfy increasing but uncertain demand. As above, the demand increases by x with probability p or it stays the same with probability $1 - p$. Let T denote the minimum number of years until the current excess capacity runs out; thus, the current excess capacity is Tx.

The firm has two options for investment. It can either invest in an expansion project with a long lead-time L_l or a short lead-time L_s. To compare these two investment options, we assume that $T \geq L_l > L_s$. Let K_l and K_s denote the costs of expansions with long and short lead-times respectively. Then, the total expected cost of these expansions are obtained as follows:

$$\begin{aligned} I_i &= \sum_{k=0}^{\infty} \overbrace{\binom{k+T-L_i-1}{T-L_i-1} p^{T-L_i}(1-p)^k}^{\text{probability of initiating construction in year } k+T-L_i} \underbrace{\left[\frac{K_i}{(1+r)^{(k+T-L_i)}}\right]}_{\text{discounted cost in year } k+T-L_i} \\ &= K_i \left(\frac{1}{1+\rho}\right)^{T-L_i} \end{aligned}$$

where $i \in \{l, s\}$ and $\rho = r/p$. Clearly, as the lead-time decreases, the expected present value of the cost decreases since it can be deferred further into the future. The value in shorter lead time can be

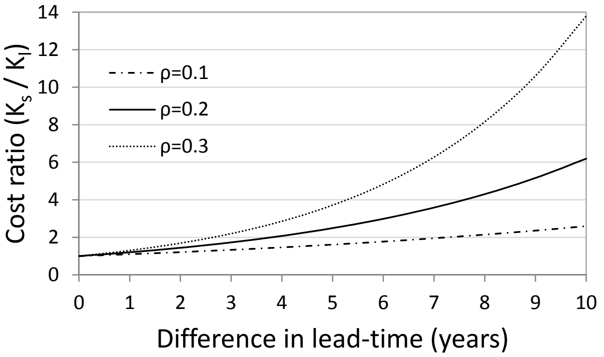

Figure 3: *The figure shows the ratio of the immediate investment cost of the short lead-time option to the long lead-time option for* $\rho = r/p = \{0.1, 0.2, 0.3\}$*, for which their expected costs are equal.*

visualized in a manner similar to the previous example. Equating the expected present value of the cost for the two lead-times, i.e. $I_l = I_s$ we see that

$$\frac{K_s}{K_l} = (1+\rho)^{L_l - L_s}$$

A difference in lead-time of only a couple of years can, for reasonable values of ρ, therefore compensate for significant increases in capital cost. The effect is illustrated in Figure 3. For example, at $\rho = 0.2$ a difference in lead-time of four years can make up for a factor of two increase in the cost of the short lead-time technology. The attractiveness of the shorter lead time scenario is obviously compounded at greater values of ρ. The reasons are the same as in the preceding section; increasing the discount rate, r, lowers the present value of future costs and decreasing the probability p of a demand increase has the consequence of further deferring these costs in the future.

4.2.3 Investment advantages of shorter lifetimes

Another feature of small unit scale technology is that it tends to have shorter lifetimes. While unit scale does not necessarily determine unit durability, large unit scale tends to be correlated with a long operating life. This is true for several reasons. Building large unit scale technology requires substantial structural elements to accommodate large weights, heights and lengths of components. Because these structural components are inherently long-lived, this tends to drive engineers to make the other system components equally long-lived, since the incremental investment needed to make the entire system long-lived is low. Also, because large unit scale projects are often produced on a customized basis, engineering the technology for long lifetimes is necessary to amortize the fixed design and construction costs.

The reverse is true for small unit scale technology; since there are no large scale support structures with inherently long lifetimes, there is no need to engineer other system components for long lifetimes. Additionally, when small unit scale technology is mass manufactured, the fixed costs of design and production capital can be amortized over large numbers of units rather than making any given unit long-lived. Thus, the choice of unit scale often has important consequences for the length of the investment cycle, which in turn can significantly impact investment returns. Specifically, when future profits are uncertain, a shorter investment cycle allows for the option of disengaging sooner if conditions turn unfavorable while retaining the option of reinvestment at a later stage.

We will here consider a simple example that serves to illustrate the value of capital lifetime. Consider a firm that has the ability to invest, with no lead-time, in a physical asset with a lifetime of n years at a cost of K_n. Regardless of lifetime, the asset produces output at a constant rate with an uncertain annual operational profit, $\pi(t)$. The profit, $\pi(t)$, can assume one of two states; a high state $\pi_{\text{high}} = 1$, and a low state, $\pi_{\text{low}} = 0$. The state of a period is observed prior to the start of the period.

The profit changes state from year to year with a probability p. The discrete time Markov process, $\pi(t)$, that characterizes the profit stream over time can be described by the transition matrix M,

$$M = \begin{array}{c} \begin{array}{cc} \pi_{\text{high}} & \pi_{\text{low}} \end{array} \\ \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \end{array}.$$

The strategy of the firm is to reinvest in capital of the same lifetime (and cost) after the current investment has expired, as soon as the observed profit reaches the high state. When considering such a strategy over an infinite time-horizon, note that the profit stream is independent of the lifetime of the individual investment. Since we have assumed zero lead-time, the firm will recommit once profits are positive, meaning that the firm is only idle when profits are zero. With a shorter lifetime of the investment, the firm can defer investment costs more frequently and more easily avoid being invested during non-profitable years.

Denote by I_n the expected present value of all future investment costs in a year when an investment is made. Since the underlying process is time-invariant and since I_n is calculated over an infinite time-horizon, the value of I_n is the same every year an investment is made. Letting ν_n be the random time between two subsequent investments, we then have

$$I_n \quad = \quad K_n + I_n \mathbb{E}_{\nu_n} \left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1 \right], \tag{11}$$

where r is the discount rate. The random time, ν_n, is the stopping time $\nu_n = \inf_t \{t \geq n; \pi(t) = 1 | \pi(0) = 1\}$, that is, the first time profits reaches the high state after n years. As shown in appendix §C, (11) can be solved for K_n:

$$K_n = I_n f_n,$$

where f_n is a function of p, r and n. Two scenarios, with capital of different lifetimes, n_1 and n_2, would be equally attractive if $I_{n_1} = I_{n_2}$, since their respective profit streams are identical. Thus, we can find the relation between their respective individual investment costs, K_{n_1} and K_{n_2}, from

$$\frac{K_{n_1}}{K_{n_2}} = \frac{f_{n_1}}{f_{n_2}}.$$

The time value of money plays a natural role in considerations regarding lifetime; with a higher discount rate, it becomes more favorable to defer investments, which would favor a shorter lifetime in relative terms. As can be seen in Fig 4, uncertainty over future profits will further increase the attractiveness of a shorter investment relative a longer lived one. For instance, assuming that there is a 10% chance that profits will change from year to year, the value of an investment lasting only 5 years is roughly half of one lasting 30 years.

4.3 Operating flexibility

Small unit scale provides increased flexibility in terms of deploying partial capacity since it is possible to selectively run varying numbers of smaller units in order to achieve a targeted level of total output. Facilities consisting of a single large unit often have to be operated in an effectively binary, "on-off" mode, producing either nothing or at maximum capacity. A coal-fired power plant, for example, has a limited range of outputs for which it can operate efficiently. As a consequence, these plants are limited to providing steady, base-load power and cannot effectively serve variable peak-load demands or efficiently slow down to avoid waste during period of low demand.

To illustrate this idea, we rely on a model of a plant consisting of multiple units operated in an on-off fashion with total capacity C. The plant has to satisfy a random demand D. Let n denote the number of units in the plant, so $C_u = C/n$ denotes the unit capacity of the equipment in the plant. The excess operational capacity, X_C, can be written as

$$X_C = kC_u - D, \quad \text{where} \quad k = \left\lceil \frac{D}{C_u} \right\rceil.$$

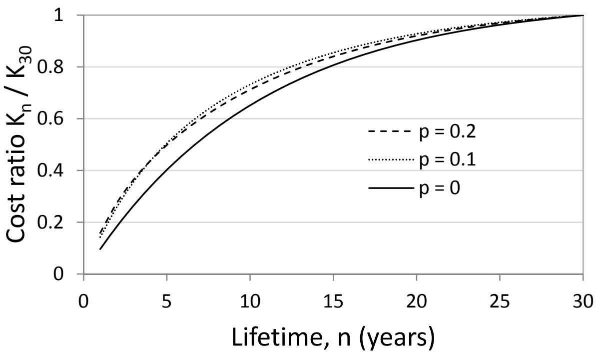

Figure 4: *The ratio of the individual investment cost, K_n to K_{30}, such that the expected present values of all future investment costs over an infinite time horizon are the same for a scenarios where the individual lifetime is n and 30 years respectively. The line $p = 0$ represents the deterministic case when there is zero time between subsequent investments. The discount rate is $r = 10\%$.*

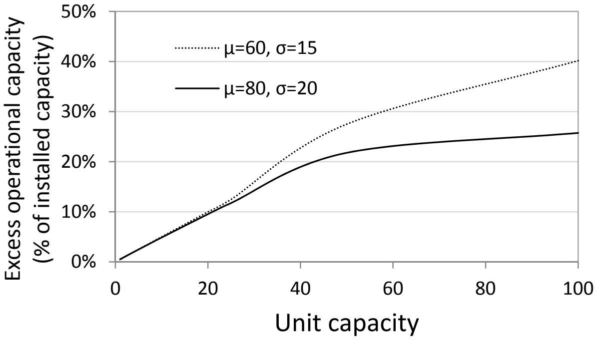

Figure 5: *The figure shows the average excess operational capacity for different unit capacities where D is a normally distributed random variable with mean μ and standard deviation σ, truncated between 0 and 100.*

The expected excess operational capacity, $\mathbb{E}[X_C]$ is

$$\mathbb{E}[X_C] = C_u\mathbb{E}[k] - \mathbb{E}[D] = C_u \sum_{k=1}^{n} kP\left[(k-1)C_u < D \leq kC_u\right] - \mu,$$

where μ is the mean demand.

Figure 5 depicts the average excess operational capacity for different unit capacities C_u using two different normal distributions for D, defined by μ and σ, both truncated between 0 and 100. It can be seen from the figure that when the unit capacity is small, it is easier to satisfy the demand with little excess operational capacity since the plant's output can be adjusted in a flexible fashion. However, as the unit capacity increases, that flexibility is lost. For example, when the unit capacity is 100 there will only be one unit in the plant, and it will have to operate all the time whether the demand is 1 or 100. Another important factor to take into account is how far the mean demand is from the maximum possible demand. As the mean demand decreases, the impact of smaller unit size on excess capacity in operation increases especially for larger unit sizes.

In the above model we implicitly assumed that storing the output of the plant under consideration is too costly or not possible. With the possibility of storing the plant's output, it is possible to achieve high utilization with large unit sizes by carrying inventory. For example, the plant may be turned on

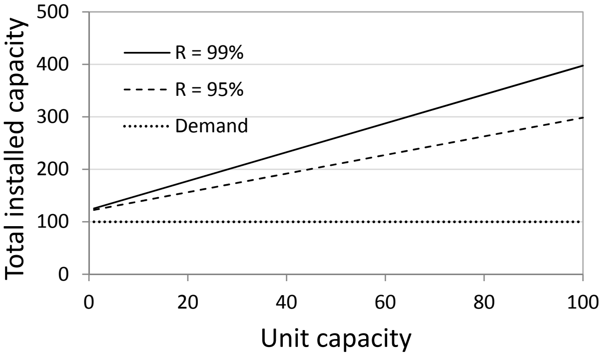

Figure 6: *The figure shows the total capacity investment required for different unit capacities and individual unit reliability* $p = 0.9$.

at full capacity when the inventory drops down to a certain threshold and can be kept in operation until the inventory reaches a target level.

4.4 Diversification

Small unit scale also provides significant diversification benefits. By exploiting statistical independence of many small operating units rather than relying on a few large operating units, it is possible to have lower unit reliability yet still achieve higher overall system reliability. If all output stems from one single large unit, a single failure can reduce output to zero, which makes it necessary to incur the costs of substantial redundancies. If, however, the same total output is provided by 10,000 units that are easily replaced, the impetus for built-in redundancies in any given unit are diminished. This reduced need for high unit-level reliability can both reduce capital costs and improve service reliability.

We illustrate this concept with the following simple example. Suppose a utility is to provide an output capacity D, available with a probability R in a given time period ($1 - R$ is the probability of failure). We assume that a single unit, with capacity C_u, can either be fully functioning with probability $p < R$ in a given time period or not at all. As a result, the utility will need redundancy to make up for the missing reliability. So, the utility has to decide on the minimum number of units to install, n^*, to ensure that the aggregate available capacity $C(n) = nC_u$ exceeds demand D with a probability of at least R. With the assumed independence of the individual units this problem can be formulated as

$$n^* = \min\left\{ n \geq \left\lceil \frac{D}{C_u} \right\rceil \middle| P(C(n) \geq D) \geq R \right\}, \quad \text{where} \quad P(C(n) \geq D) = \sum_{s=\left\lceil \frac{D}{C_u} \right\rceil}^{n} \binom{n}{s} p^s (1-p)^{(n-s)}.$$

For different R and C_u, Figure 6 shows the total capacity the utility has to invest in when $D = 100$ and $p = 0.9$. From the figure we can infer that as the size of the individual units increase, the amount of excess capacity that needs to be installed significantly increases. For example, when $R = 0.95$, the amount of excess capacity needed is five times more when the unit size is 50 compared to a unit size of 1. As the system reliability increases, this difference also increases, and the excess capacity for a unit size of 50 becomes 7-8 times the amount needed for a unit size of 1 in the case of $R = 0.99$.

4.5 Tipping points

Considering the factors discussed above, firms make a choice of unit scale based on the combined effect of these various costs and benefits. The resulting total cost as a function of unit scale can behave in one of two ways. There can be an intermediate unit scale that is optimal, with total costs

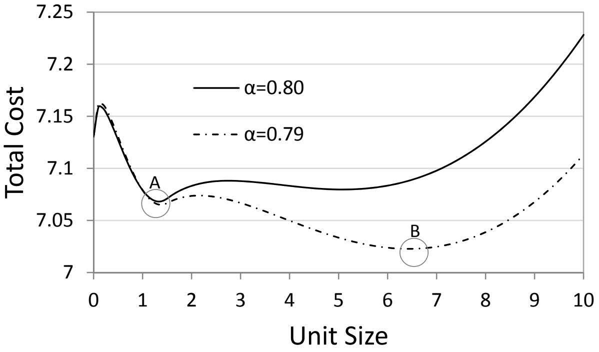

Figure 7: *The figure shows the total investment cost* (12) *as a function of unit size for different scale parameters.*

increasing as one deviates either up or down from this optimal point. Alternatively, total costs can exhibit "bang-bang" behavior, with costs either increasing or decreasing with unit scale, resulting in an optimum that occurs either at the smallest or largest practical unit scale. In both cases, changes in cost functions, scaling relationships or scale-dependent flexibility and diversification benefits – such as those driven by advances in technology – can lead to a "tipping point," at which the optimal scale switches discontinuously from large to small.

In the case of an intermediate optimal scale, a tipping point may occur if the cost function has multiple local minima and if the global minimum shifts from one local minimum to another as model parameters change. For example, consider the following single-period model where a firm decides how much capacity C to invest in order to cover a single period random demand D. For each unit of unsatisfied demand, the firm incurs a penalty of q. Capacity costs exhibit economies of scale, modeled as a power function. Then, the total cost, $K(C)$, is

$$K(C) = \underbrace{k_0 C^\alpha}_{\text{capacity cost}} + \underbrace{q\mathbb{E}\left[\max\{0, D - C\}\right]}_{\text{penalty cost}}, \tag{12}$$

Figure 7 shows the total cost as a function of unit size for two different scale parameters ($\alpha = 0.79$ and $\alpha = 0.80$) when $K_0 = 0.8$, $q = 1$ and D follows a piecewise uniform distribution with the following probability density function

$$f(D) = \begin{cases} 0 & \text{if } D < 0 \text{ or } D > 23, \\ 0.297 & \text{if } D \in [0, 1.5], \\ 0.026 & \text{if } D \in (1.5, 23]. \end{cases}$$

Note that in Figure 7 there are two locally optimal unit sizes for both values of α, and 0.80 is a tipping point, since at that value the optimum at the larger unit capacity B no longer has the lowest cost but instead the local optimum A at a smaller unit capacity has the lowest cost. While this is a simplified example, it serves to illustrate that even when capacity costs exhibit economies of scale, a small shift in the underlying technology- or scale parameter- can radically impact the optimal unit scale.

In other cases, cost functions may exhibit bang-bang behavior. Consider the following model where we need to invest in $C_{\text{tot}} = 100$ units of capacity and have the option to invest in different unit scales. Assuming the relationships described in sections 3.1.1 and 3.1.2 hold, we obtain the following formula for capital cost:

$$K(C) = k_0 C^\alpha N^{\log_2 \varepsilon + 1} = k_0 C^\alpha \left(\frac{C_{\text{tot}}}{C}\right)^{\log_2 \varepsilon + 1} \propto C^{\alpha - 1 - \log_2 \varepsilon}, \tag{13}$$

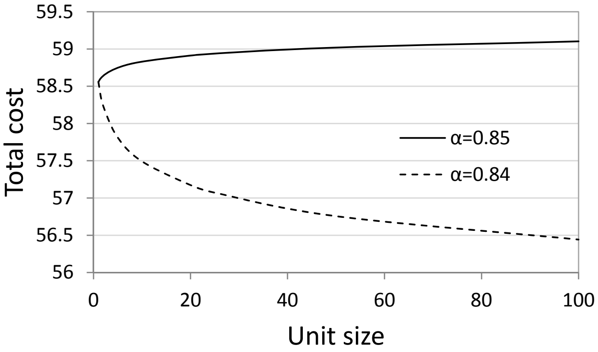

Figure 8: *The figure shows the total investment cost* (13) *as a function of unit size for different scale parameters ($\varepsilon = 0.9$ and $K_0 = 1$).*

where C again denotes the unit size of the investment and $N = C_{\text{tot}}/C$ is the number of units installed. For different scale parameters, Figure 8 shows the total cost of investment as a function of the different unit scales. In this example, the tipping point for α is about 0.85. When $\alpha \geq 0.85$, the total cost function switches from being decreasing to increasing in unit scale, and hence a radical shift in the optimal unit scale occurs.

The evolution of super computer technology illustrates a powerful real-world example of this basic tipping point phenomenon. As noted previously, the traditional approach to achieving high capacity and greater speed in the early life of an industry was to build increasingly powerful machines with greater and greater processing speed and also increasingly large size. The 'supercomputer market crash' of the mid 1990s occurred when mass-produced CPUs from the expanding personal computer industry reached a tipping point. Somewhat suddenly, it became less costly to achieve high computing capacity through large numbers of small-scale CPUs in parallel as we see in computer technology today, and demand for traditional super computers collapsed. This historical example underscores the point that sudden shifts in optimal unit scale due to technological advances can occur suddenly and result in dramatic disruptions of entire industries.

5 Technologies on the verge of a tipping point toward small unit scale

We next look at three examples of technologies that appear ripe for a shift to radically smaller unit scale. While every technology must be evaluated on a case-by-case basis, these three technologies have evolved to large unit scale without clear physical benefit and therefore serve as a natural starting point. While anecdotal, these examples point to the potential broader benefit of re-examining the aging orthodoxy of bigger-is-better in other infrastructure industries.

5.1 Ammonia synthesis

Among the most economically important catalytic processes to date is ammonia synthesis. Since its inception almost a century ago, the synthesis of ammonia through the Haber-Bosch process has provided the world with synthetic fertilizers, thereby dramatically altering Malthusian projections of population levels constrained by food supply. Except for minor alterations to the catalyst, the process looks very much the same throughout the $50 billion dollar industry of today [31, 34] as it did during the early days of commercial implementation. One feature, however, has dramatically changed: unit size. The first BASF plant had a capacity of 30 metric tons per day (MTPD) [45]. In comparison, the currently planned urea facility in Collie, Australia has a nominal capacity of 3,500 MTPD, making it the largest in the world [73]. An increase in unit capacity by two orders of magnitudes epitomizes the aforementioned trend in the evolution of unit size.

The synthesis of ammonia, as with most other catalytic processes, requires careful control of temperatures throughout the reactor in order to maintain favorable kinetics, thermodynamic equilibria conditions and stability of the catalyst. Managing the temperature profile throughout a reactor in this process becomes progressively more difficult when increasing the reactor size; indeed, modern ammonia synthesis reactors require additional internal heat exchangers to maintain optimal process parameters. Additionally, ammonia synthesis typically occurs at a total pressure of around 200 atmospheres. At such high pressures, the mechanical potential energy of a pressurized reactor of standard size is on the GJ scale, which in itself poses substantial safety concerns. The fact that the gases involved are highly explosive further exacerbates the consequences of a critical failure.

Assuming sufficient capabilities of automation such that labor costs can be decoupled from the number of individual units deployed, a decrease in individual reactor size could potentially reduce overall costs. First, decreasing unit size to the point that no internal heat sinks are necessary would reduce the complexity of the individual reactor[3]. Also, the wall thickness of a pressure vessel scales with the linear dimension of the reactor at constant pressure, and the reactor material will scale proportionally to volume and hence also throughput. Therefore, the total amount of material used will, to a first approximation, remain constant or decrease when deploying multiple smaller units with the same aggregate capacity of a typical ammonia synthesis reactor. Furthermore, an array of parallel smaller units would substantially mitigate the impact of catastrophic failure of a single reactor. Also, catalytic processes, including the synthesis of ammonia, all suffer the problem of catalyst deactivation. In a monolithic setting, regenerating or replacing catalysts requires a complete albeit temporary shutdown of the reactor. Referring to the operational flexibility arguments, a more modular plant consisting of parallel units would be less exposed to the risk of such complete but unavoidable outages. Instead of bringing the entire output to a standstill, individual units could be swapped out and repaired off-line. These factors all point to potential cost savings with a modular infrastructure approach.

In the case of ammonia synthesis, there are two other major process steps upstream of the main synthesis step: nitrogen and hydrogen production. These two sub-processes, typically cryogenic distillation of air and steam-methane reforming, have evolved along with the conversion process to large unit sizes. If the ammonia conversion were to occur at substantially smaller unit sizes, the size of the upstream processes should also be reevaluated, perhaps to the benefit of other methods of producing the constituents of the Haber-Bosch process, e.g. membrane air separation technologies and water electrolysis. It is worthwhile mentioning that the main use of ammonia is in fertilizers. Serving mainly the agriculture industry, the demand for the end-product is extremely diffuse, suggesting potential benefits of distributed operation. Additionally, if water can be used as a source of hydrogen through electrolysis, both inputs and outputs of the process are ambient, further strengthening the case for small-scale, and distributed ammonia production.

5.2 Water desalination

One of the main engineering challenges of the coming century is to create large, stable and affordable supplies of fresh water from the earth's largest water reservoir, the oceans [62]. Among the various desalination technologies available for this purpose, reverse osmosis (RO) has seen the largest growth in recent years and represents almost half of the $20 billion desalination market today [36, 35, 30]. Some regions, notably the Middle East, parts of Australia and several island communities around the world, have come to rely on reverse osmosis desalination as a base load source of fresh water.

Defining a unit scale in RO operations is less straightforward than in the previous examples because the membranes are currently manufactured and assembled in small units called modules. Regardless, the core process of separation in a modern RO-plant encompasses three components: a

---

[3] Performing a crude heat balance analysis, assuming uniform temperatures across the reactor with otherwise typical reactor parameters (T = 700K, catalyst activity = 10μmol/g·s), a heat transfer coefficient through the reactor walls of 10W/m^{2}K (heat absorbing medium is air without forced convection [23]) and a reaction enthalpy ΔH = $-$50kJ/mol NH$_3$ reveals that auto-thermal conditions (generated process heat is balanced by losses to ambient air) apply with a cylindrical radius of about 0.5 cm.

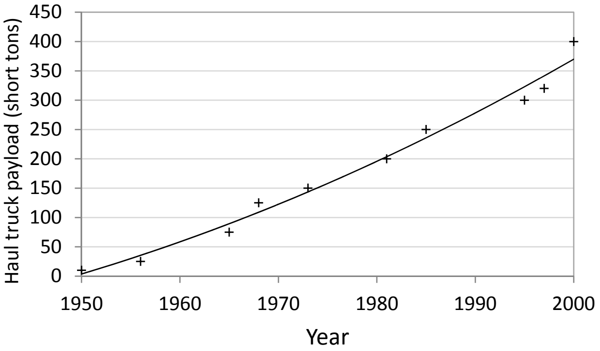

Figure 9: *Evolution of the payload of mining haul trucks. Data adapted from [48].*

high-pressure pumping system, membrane modules housed in parallel pressure vessels, and an energy recovery system. One of the most modern and efficient RO desalination plants in the world is the Ashkelon plant in Israel with a capacity of 330,000 cubic meters of fresh water per day [66]. With eight high-pressure pumps, the 40,000 cubic meters of fresh water per day and per pumping system can serve as a benchmark for unit size in current RO desalination.

Almost half the cost of desalinated seawater through RO can be attributed to energy. Of the remaining cost components, capital costs dominate, leaving only a minor fraction to other variable costs [79]. With such a cost structure, the specific energy requirement (energy required per unit output) serves as a decent indicator of the total price of a given RO desalination operation. Conventional wisdom would suggest that to increase the physical efficiency of a process like RO, and hence lower the specific energy consumption, the strategy would be to scale up unit size. Indeed, the aforementioned plant in Ashkelon is not only one of the largest plants in the world but also one of the most efficient with electricity requirements of 3-4 kWh/m^3 of produced water [66]. The notion that increased energy efficiency in RO has to be accompanied by increased unit size is, however, contradicted by examining desalination systems found on recreational boats. With a capacity of little more than one cubic meter per day, these small, modular systems consume only 3.8 kWh/m^3 [77], which is on par with the utility-sized operation. While based on the exact same technology, the tight on-board space constraints have resulted in process designs that have similar levels of energy consumption but with smaller footprints. This indicates that a truly modular design in RO desalination could lead to additional future cost savings as production of small units is scaled up to match large scale demand.

5.3 Mining

The value of U.S. domestic production of raw materials from mining was estimated at $64 billion dollars in 2010 [13]. Including further downstream processing, the raw materials sector accounts for a substantial part of GDP and is the foundation of industrial economies. While mining operations differ substantially for different minerals in varying geologic formations, the task of hauling ore from the point of excavation to the initial processing site is worth examining from the perspective of unit scale. We focus here on operations in open pit, or surface mining, but the concepts apply more broadly to other mining operations.

Labor productivity is a key metric in evaluating mining operations because labor accounts for a large fraction of total costs. Therefore the most natural way to increase profitability of a given mine has been to scale up the size of individual process equipment such as loaders and haulers. Indeed, as seen in Figure 9, the size of the largest available haulers has increased by a factor ten over the past 50 years. A general consequence of this trend is that smaller mines, which preclude the use of larger equipment [25], become less profitable, and hence mining operations tend to be more concentrated on large mines, as documented by [42] in the case of the global copper industry from 1975 to 2000.

In a report on the future of the mining industry published by the Rand corporation in 2001

[64], the opinion among key industry insiders was divided as to whether there would be a continued increase in truck sizes or if these economies of scale had reached a peak. In hindsight, knowing that truck sizes have indeed stagnated and the very largest trucks have stalled in the 400 ton class (as can be observed by considering the current portfolio of Caterpillar), the latter viewpoint seems to have prevailed. Some of the possible factors behind this apparent trend are presented in [25]. For instance, auxiliary civil works, e.g. roads and bridges, to accommodate larger trucks going in and out of a mine become more costly. Moreover, larger equipment diminishes the possibility of selective mining techniques, thus resulting in the transportation of lower grade ores for further processing. The complexity of larger machinery also increases markedly at the largest end of the spectrum and hence requires additional training of operators and repair crews as well and larger (and more expensive) maintenance facilities.

With these issues in mind as well as considering workers' safety, automating various mining processes is a potentially attractive means for future cost reductions. Additionally, with mining typically being a remote operation, additional infrastructure has to be provided in order to accommodate on-site labor. According to [22] the cost of one mining truck driver in remote areas of Australia amounts to $150,000 per year[4], of which more than $36,000 goes to auxiliary support such as transportation, accommodation and food. Operating in three shifts, this translates into $450,000 per year per truck in labor costs, which does not include personnel in training. While truck prices are hard to ascertain exactly, assuming that the investment required is on the order of $5 million, the capital charges at 10% interest are almost on par with labor costs, making the potential benefits of automation apparent.

Even though tests have been performed recently on operating retro-fitted autonomous mining trucks [22] in Australian mines, such technology has not yet caught on. With non-stationary and interacting robotic systems making progress by the day, as manifested by Google's autonomous car [16] and 'Junior', the driverless vehicle developed by Volkswagen and Stanford through DARPA [69], it is only a matter of time before such technology becomes viable in isolated areas such as a mine. When the technology does become available, there is little reason automation should proceed with the ultra-large-size equipment of today and not with much smaller units, perhaps in the 1-10 ton class. In addition to the flexibility arguments raised in previous examples favoring small unit size, smaller automated units can make smaller mines economical alongside large ones, thus increasing the total resource base. Also, from a physical perspective, the dead weight to payload ratio is likely to decrease with smaller trucks, suggesting potential fuel savings as well.

6 Conclusion: Learning to "think small"

When choosing from a palette of available technologies, the ultimate decision has historically been predicated on a positive response to the question: Does the technology "scale up?" Yet as we have argued in this paper, our increased ability to automate and control processes without the presence of either the human hand or mind, the capability of mass production to drastically reduce capital costs, and a more enlightened view of the total costs and benefits of unit scale, casts significant doubt on the validity of the bigger-is-better criterion. As we have argued, scaling up in numbers – rather than in unit size – can provide many of the same benefits of large unit scale and offers entirely new benefits that can only be achieved with small unit scale. Consequently, in our view, the fundamental decision processes surrounding the choice of technologies and their implementation need to be revisited.

Doing so, however, requires an entirely new mind set; educators, engineers, business leaders, financiers, standards bodies, regulators – the entire industrial ecosystem – must learn to "think small." In order to reap the benefits of small unit scale and achieve the needed paradigm shift, institutional biases towards large-scale must be eliminated and knowledge about how to think small must be developed.

Engineers, for one, need revised training and new conceptual tools. In today's engineering schools, students are instilled with the notion that unit scale-up is a precondition for the viability of most technologies. So consequently, they focus on designing for scale economy. Instead, they must learn how to design small – *design for granularity* as it were. Small modular technologies designed to

---

[4] These numbers are given in $ Australian but with an exchange rate of roughly 1:1 the numbers are almost equivalent to $ U.S.

function in massively parallel configurations should not look like miniature versions of behemoth industrial plants; they require their own distinct approach to design – one that emphasizes off-loading control functions to central controllers, simplifies functionality to minimize the need for ongoing control and maintenance, and reduces part counts by creating more integrated components. Engineering small requires designs that aim to leverage the economies of mass production and exploit the power of automation and sensors to eliminate the need for human labor. Only by applying such design principles can the full benefits of small unit size be realized. Engineers must also be exposed to examples of designs that are optimized for granularity, so they can develop an instinct for how it is done.

Business leaders and financiers must likewise revise their approaches to project evaluation. Long-standing net present value (NPV) evaluations based on simplistic pro-forma projections of factors such as demand, cost, price, unit reliability and so on, must be replaced by the use of more sophisticated models (of the sort illustrated in our paper) that accurately account for the many inherent flexibility and option-value benefits of small unit scale. Only then will their decisions be driven by the true economic costs and benefits of unit scale.

Lastly, a mass market for small unit scale technology will not flourish unless industry leaders recognize the potential of thinking small and create the necessary standards and common interfaces needed to open their markets to small modular technology. Indeed, one of the reasons behind the astonishing developments of the computer industry in the past century was the abandonment of the mentality that every component had to be manufactured in-house. Opening up the black box that was the computer to outside parties allowed firms to focus on fewer parts and also forced the industry to adopt standards, resulting in the "plug-and-play" environment that eventually made the PC possible and so dramatically successful. In addition, down-sizing, standardizing and then proliferating technologies to a larger domain of applications further reduces costs by increasing the aggregate market size for each technology. It also increases the likelihood of applications not yet thought of; after all, the early pioneers of the computing industry could hardly have anticipated the multitude of applications that permeate virtually every part of our society today, like smart phones and video games.

These changes in mindset and industry norms will take time to develop; massively parallel plants will not suddenly appear overnight. Indeed, there is considerable inertia in most industries that may impede the transition to thinking small for many years to come. One explanation for this inertia is known as the "lock-in effect" in the economics literature. As in the case of the QWERTY keyboard, once a technology establishes dominance early on, a later superior technology may not be able to gain market share [28]. In his seminal work, Arthur [21] shows that this kind of a behavior is observed in industries with increasing returns to learning, as is the case with large-scale infrastructure. But two factors offer hope that this lock-in can be overcome. For one, as demonstrated in the preceding sections, once the flexibility and diversification benefits of small-scale technology are recognized, these may tip the scales toward adoption. Secondly, niche applications of small-scale technologies in areas where larger scales are infeasible or too costly may allow firms to accumulate the necessary experience to compete with large-scale technologies in conventional markets, as was the case with microcomputers. Once this transition happens and small scale thinking takes root, it has the potential to radically disrupt entire industries. Like the behemoth reptiles of the Cretaceous period, individuals and firms caught on the wrong side of such a meteoric transition will likely suffer.

Yet despite the great promise of thinking small, we are not arguing that small-scale technology is a panacea. Indeed, some enterprises adopting a small-scale strategy have had failures such as Raser Technologies[5]. And there will always be a role for large unit scale; massive rivers require massive hydroelectric dams, after all. Still, the concept that every industrial process with large aggregate output *requires* large unit scale technology to match is fundamentally flawed and inherently limiting.

---

[5] Raser Technologies was a Provo, Utah based geothermal-energy developer that sought bankruptcy protection in 2011 [57]. It focused on constructing geothermal power plants that used small modular equipment and had only one plant in operation, which was located in Beaver, Utah, when it declared bankruptcy. The reasons behind its failure are disputed. Raser sued its technology provider Pratt & Whitney Power Systems for fraud regarding the efficiency of the units [61]. Pratt & Whitney on the other hand claimed that the site of the power plant was not suitable [53]. Raser's shareholders also filed a class action lawsuit against some of its current and former directors, and officers over failing to disclose material adverse facts regarding the power plant's site [56].

We will all benefit from a more enlightened world in which the ability to “scale up” does not dictate our choice of technology.

Appendices

A Scaling laws for elastic bodies

Applying the theory of continuum mechanics to the scaling of engineered structures allows for a number of conclusions regarding the prospects of scaling up a solid body in size. The solid body can be characterized by its material distribution and once this body is subjected to forces, both body forces like gravity and contact forces like pressure on surfaces, it will deform, causing internal stresses throughout the material. Most engineered structures are designed to operate within the elastic limit, i.e. when loads are removed the structure will revert to its original shape. In this elastic region, internal stresses and material displacements, the latter described by a strain tensor, typically exhibit a linear relationship. The equations governing the dynamics of a solid body can be written as

$$\begin{aligned} \partial_i \sigma_{ij} + F_j &= \rho \partial_t^2 u_j, && (14)\\ \sigma_{ij} &= C_{ijkl} \varepsilon_{kl}, && (15)\\ \varepsilon_{ij} &= \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right), && (16) \end{aligned}$$

where F denotes the body forces, σ is the stress tensor, ϵ is the strain tensor and where C is the fourth order elasticity tensor that describes properties of the material, all being functions of the spatial location x as well as of time. The function $u(x,t)$ describes the displacement of a mass particle, originally at x_0, at time t. That is, $x = u(x,t) + x_0$. The equations (14)-(16) are defined on $\mathcal{U} \times \mathbb{R} \subset \mathbb{R}^4$, with boundary $\partial\mathcal{U}$. To completely determine the problem, boundary conditions and initial conditions have to be specified. The boundary conditions are typically of mixed type, where traction T is specified on one part of the boundary, $\partial\mathcal{U}^T$, and displacements are given on the complementary part, $\partial\mathcal{U}^u$, with $\partial\mathcal{U} = \partial\mathcal{U}^T \cup \partial\mathcal{U}^u$.

$$\left\{ \begin{array}{lcll} \sigma n &=& T, & x \in \partial\mathcal{U}^T, \\ u &=& \overline{u}, & x \in \partial\mathcal{U}^u, \end{array} \right. \qquad (17)$$

$$\left\{ \begin{array}{lcll} u(x,0) &=& u_0(x), & x \in \mathcal{U} \\ \partial_t u(x,0) &=& \dot{u}_0(x), & x \in \mathcal{U} \end{array} \right. \qquad (18)$$

By rescaling the problem we intend to study a similar body where all spatial dimensions have been scaled by a factor $\lambda > 0$ and time has been scaled by a factor $\mu > 0$. This rescaling differs from a mere coordinate transform in that we are interested in keeping the boundary forces constant. For example, the pressure on a submarine at a given depth is approximately independent of the size of the vessel. The scaled domain is denoted by $\mathcal{U}_{\lambda,\mu}$. That is, for any function $\tilde{f}$ defined on $\mathcal{U}_{\lambda,\mu}$ we have

$$\tilde{f}(\xi,\tau) = \tilde{f}(\lambda x, \mu t), \qquad (19)$$

where $(\xi,\tau) \in \mathcal{U}_{\lambda,\mu}$ and $(x,t) \in \mathcal{U}$. Furthermore, we say that $\tilde{f}$ on $\mathcal{U}_{\lambda,\mu}$ is generated by a function f on $\mathcal{U}$ if

$$\tilde{f}(\xi,\tau) = k f(x,t), \qquad (20)$$

where k is some constant. If $k = 1$ we call $\tilde{f}$ on $\mathcal{U}_{\lambda,\mu}$ unitarily generated by f on $\mathcal{U}$. From (19) and (20) it follows that the spatial derivatives of a generated function evaluate to

$$\partial_i \tilde{f}|_{(\xi,\tau)} = \frac{k}{\lambda} \partial_i f|_{(x,t)}, \quad \text{or simpler,} \quad \partial_i \tilde{f} = \frac{k}{\lambda} \partial_i f, \qquad (21)$$

and for time

$$\partial_t \tilde{f} = \frac{k}{\mu} \partial_t f. \qquad (22)$$

Solution to scaled problem

We will investigate possible solutions to the scaled problem on $\mathcal{U}_{\lambda,\mu}$, that is

$$\begin{aligned}\partial_i\tilde{\sigma}_{ij} + \tilde{F}_j &= \tilde{\rho}\partial_t^2\tilde{u}_j, & (23)\\ \tilde{\sigma}_{ij} &= \tilde{C}_{ijkl}\tilde{\varepsilon}_{kl}, & (24)\\ \tilde{\varepsilon}_{ij} &= \frac{1}{2}\left(\partial_i\tilde{u}_j + \partial_j\tilde{u}_i\right). & (25)\end{aligned}$$

The magnitude of the boundary forces is assumed to be the same as in the initial problem, i.e. $|\tilde{\sigma}n| = |\tilde{T}| = |T|$. Given a solution $u(x,t)$, together with its derived entities $\varepsilon(x,t)$ and $\sigma(x,t)$, to the initial problem (14-18) defined on $\mathcal{U}$ we will show that for certain choices of the scaling parameters (λ, μ) the generated function $\tilde{u}(\xi,\tau) = \lambda u(x,t)$ solves the scaled problem for a family of generated functions

$$\begin{aligned}\tilde{F}(\xi,\tau) &= \alpha F(x,t),\\ \tilde{\rho}(\xi,\tau) &= \beta\rho(x,t),\\ \tilde{C}(\xi,\tau) &= \gamma C(x,t),\end{aligned} \qquad (26)$$

for a certain parameter ensemble (α, β, γ). The choice of parameter ensemble has important physical interpretations. For instance, if $\beta = 1/2$ the material in the scaled problem has been selected to have half the density of the original material. Note that our choice of $\tilde{u}(\xi,\tau) = \lambda u(x,t)$, where λ is the spatial scaling coefficient, leads to the same strain as in the original problem, regardless of the choice of parameter ensemble for the other functions. This can be seen by using (21),

$$\tilde{\varepsilon}_{ij}(\xi,\tau) = \frac{1}{2}\left(\partial_i\tilde{u}_j(\xi,\tau) + \partial_j\tilde{u}_i(\xi,\tau)\right) = \frac{1}{2}\left(\frac{\lambda}{\lambda}\partial_i u_j(x,t) + \frac{\lambda}{\lambda}\partial_j u_i(x,t)\right) = \varepsilon_{ij}(x,t). \qquad (27)$$

That is, the new strain is unitarily generated by the original strain. Assuming the same material for the scaled problem suggests that $\tilde{\rho}$ and $\tilde{C}$ on $\mathcal{U}_{\lambda,\mu}$ are both unitarily generated by their counterparts on $\mathcal{U}$, i.e. $\beta = \gamma = 1$. Then, since the strain $\tilde{\varepsilon}$ is unitarily generated, the stresses, $\tilde{\sigma}$, follow suit

$$\tilde{\sigma}_{ij}(\xi,\tau) = \tilde{C}_{ijkl}(\xi,\tau)\tilde{\varepsilon}_{kl}(\xi,\tau) = C_{ijkl}(x,t)\varepsilon_{kl}(x,t) = \sigma_{ij}(x,t). \qquad (28)$$

This means that the pressure boundary condition (17) is fulfilled. By using (22) we see that to meet the initial condition,

$$\partial_t\tilde{u}(\xi,0) = \frac{\lambda}{\mu}\partial_t u(x,0) = \frac{\lambda}{\mu}\dot{u}_0(x), \qquad (29)$$

we need $\lambda = \mu$. That is, time scales with the same factor as the spatial dimensions. Finally, the dynamics of the system read

$$\begin{aligned}\partial_i\tilde{\sigma}_{ij} + \tilde{F}_j &= \tilde{\rho}\partial_t^2\tilde{u}_j & (30)\\ \frac{1}{\lambda}\partial_i\sigma_{ij} + \alpha F_j &= \frac{1}{\lambda}\rho\partial_t^2 u_j. & (31)\end{aligned}$$

Thus, if $\alpha = 1/\lambda$, we have found a solution to the scaled problem $\mathcal{U}_{\lambda,\lambda}$ in $\tilde{u}(\xi,\tau) = \lambda u(x,t)$.

A physical interpretation of the result above is that as long as body forces decrease with the same factor as the spatial dimensions increase, a body can be scaled up and subjected to the same boundary conditions while still exhibiting the same strains. If the original body had a resonance frequency at f, the scaled version will have a resonance mode at f/λ and all other dynamics will have slowed by a factor $1/\lambda$. If body forces can indeed be scaled (for instance by altering electric and magnetic fields) the structure in question can be symmetrically scaled and operated under identical conditions. For most engineering purposes the body forces are gravitational and scaling them down is not within our reach as long as we use the same material and hence, a uniform scaling of the structure is not possible[6]. However, the solution still holds in the limit where boundary forces are of

---

[6] However a bridge or other structure built on the moon would behave like its counterpart on Earth if all dimensions are scaled up by the ratio of gravitational accelerations which is approximately 6 : 1

such a magnitude that body forces can be disregarded. This would be the case in practical situations as e.g. high-pressure vessels, submarines at depth and hauling truck beds (considering the load as a boundary condition).

Note that if either body forces can be scaled down or ignored, the amount of material used is in the scaled structure is

$$\int \tilde{\rho} d\mathcal{U}_{\lambda,\mu} = \lambda^3 \int \rho d\mathcal{U}. \tag{32}$$

For situations where body forces cannot be disregarded, a structure cannot be uniformly scaled and operated under similar conditions. If body forces are gravitational it seems likely that a scaled up structure need to be augmented in the structural elements to be able hold the increased weight. We can therefore generalize and claim that if the same materials are used in a scaled up structure, the amount required scales *at least* as λ^3. Consequently, we can dismiss the claims in the literature, see e.g. [37, 41, 74, 59], that the amount of material required when scaling up goes like the surface area, that is, like λ^2. These claims are then usually linked to the assumption that the cost of a structure is proportional to the amount of material used, and hence that cost scales accordingly.

If the body forces are gravitational, one way of achieving uniform scale-up is to select a lighter material with otherwise identical properties. While there of course are limits to such selections we show such scaling works. Again, we assume that the solution $\tilde{u}$ to the scaled problem on $\mathcal{U}_{\lambda,\mu}$ is generated from the original solution according to $\tilde{u}(\xi,\tau) = \lambda u(x,t)$. With gravitational body forces only we have, in the original problem,

$$F(x,t) = \rho(x,t)g, \tag{33}$$

where g is the acceleration of gravity which we assume constant. Hence, if we in the scaled version assume a density scaling $1/\lambda$, i.e. $\tilde{\rho}(\xi,\tau) = (1/\lambda)\rho(x,t)$, we have managed to correctly scale the body forces. However, for the dynamic equation we now have

$$\begin{aligned} \partial_i \tilde{\sigma}_{ij} + \tilde{\rho} g_j &= \tilde{\rho} \partial_t^2 \tilde{u}_j & (34) \\ \frac{1}{\lambda}\partial_i \sigma_{ij} + \frac{1}{\lambda}\rho g_j &= \frac{\lambda}{\lambda\mu^2}\rho\partial_t^2 u_j. & (35) \end{aligned}$$

By scaling the time domain by a factor $\sqrt{\lambda}$, i.e. $\mu = \sqrt{\lambda}$, the result comes out correct on the right hand side in (35). The problem is now solved if the initial velocity conditions in (18) can be scaled by a factor $\sqrt{\lambda}$ and if the dynamics are allowed to be slowed down with a factor $1/\sqrt{\lambda}$. It is straightforward to show that the same result can be found by using a material with the same density as in the original problem but where the stiffness has increased by a factor λ, i.e. $\tilde{C}(\xi,\tau) = \lambda C(x,t)$.

B Locational flexibility

We are considering the total area-dependent cost:

$$K(A) = K_T(A) + K_C(A). \tag{36}$$

Both the transportation cost, $K_T(A)$, and the capital cost per unit output $K_C(A)$ are assumed to follow power laws. For $K_T(A)$ we have

$$K_T(A) = K_T(A_0)\left(\frac{A}{A_0}\right)^{\beta},$$

where $\beta > 0$. In section §4.1, we derived an expression for the exponent in the power law for the capital cost per unit output, $K_C(A)$, influenced both by economies of unit scale and the economies of mass production. For brevity we write this power law here as:

$$K_C(A) = K_C(A_0)\left(\frac{A}{A_0}\right)^{-\gamma},$$

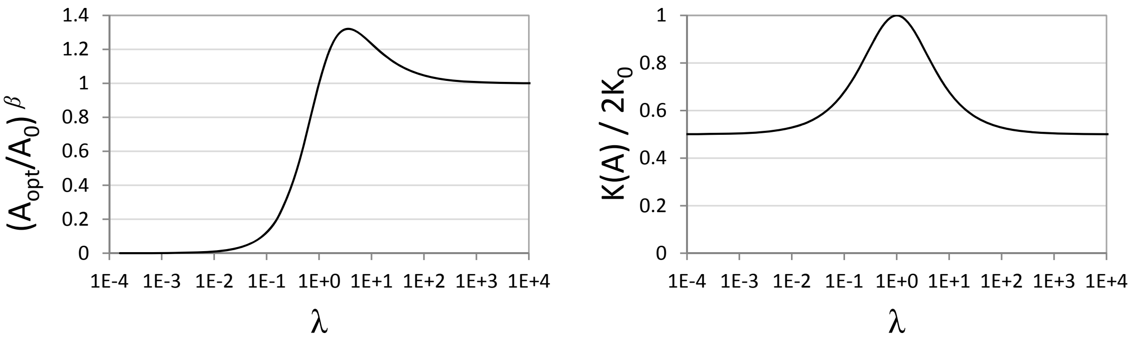

Figure 10: *Left. Implicit relation between the optimal area, A_{opt}, and the reference area, A_0, defined by the intersection of capital cost and transportation cost per unit delivered, i.e. $K_C(A_0) = K_T(A_0) = K_0$. Right. The dimensionless capital cost per unit delivered, $K(A_{opt})/2K_0$, at the optimal area. The denominator, $2K_0$, is the total cost per unit delivered at the reference area, A_0.*

where $\gamma = -(\alpha - (1 + \log_2 \varepsilon))$. From empirical observations, γ is likely to be close to zero. The case when $\gamma < 0$, i.e. when both capital and transportation costs are increasing functions of the area, results in the optimal area being as small as possible, as discussed in section §4.1. In the case when $\gamma > 0$ the total function, $K(A)$, is convex with a finite optimum, A_{opt}. Furthermore, there is a point where the transport cost, K_T, is equal to the capital cost, K_C. Choosing the reference area, A_0, as this point, i.e. where $K_C(A_0) = K_T(A_0) = K_0$, we know that total area-dependent costs are dominated by capital costs for $A < A_0$, and transportation dominated for $A > A_0$. We find the optimal area, A_{opt}, from

$$\begin{aligned} \frac{dK}{dA}(A_{\text{opt}}) &= 0 \\ \frac{A_{\text{opt}}}{A_0} &= \left(\frac{\gamma}{\beta}\right)^{\frac{1}{\beta+\gamma}}. \end{aligned}$$

From this expression we notice that

$$\left(\frac{A_{\text{opt}}}{A_0}\right)^{\beta} = \left(\frac{\gamma}{\beta}\right)^{\frac{\beta}{\beta+\gamma}} = \lambda^{\frac{1}{1+\lambda}} \tag{37}$$

where $\lambda = \gamma/\beta$. Furthermore, the total cost can also be expressed through this one parameter, λ:

$$K(A_{\text{opt}}) = K_0 \left(\left(\frac{A_{\text{opt}}}{A_0}\right)^{\beta} + \left(\frac{A_{\text{opt}}}{A_0}\right)^{-\gamma} \right) = K_0 \left(\lambda^{\frac{1}{1+\lambda}} + \lambda^{-\frac{\lambda}{1+\lambda}} \right). \tag{38}$$

These two functions are displayed in Fig 10 from which we can draw two qualitative conclusions. First, the total area-dependent cost is at best a factor 1/2 of the cost at the reference point where capital cost balances transportation cost. Second, from the implicit relation between the optimal area, A_{opt}, and the reference area, A_0, we see that for values of λ close to zero, the optimal area is order(s) of magnitude smaller than A_0 and the optimality occurs in the capital dominated regime.

C Investment flexibility

In the two examples on the benefits of increased modularity and shorter lead-time, we have made use of the following identity:

$$\frac{1}{(1-x)^m} = \sum_{k=0}^{\infty} \binom{k+m}{k} x^k. \tag{39}$$

That is, the coefficient $\binom{k+m}{k}$ is the kth order coefficient in the Taylor expansion of $\frac{1}{(1-x)^m}$ around $x = 0$.

In the example on value as a function of lifetime we are considering a discrete time, binary process $\pi(t) \in \{0, 1\}$ described by the transition matrix M,

$$M = \begin{array}{cc} \pi_{\text{high}} & \pi_{\text{low}} \\ \left(\begin{array}{c} 1-p \\ p \end{array}\right. & \left.\begin{array}{c} p \\ 1-p \end{array}\right) \end{array}.$$

The Markov property of this process admits that the present value of all future investments, I_n, is the same every year an investment is made. With the random time ν_n, between consecutive investments, we have

$$I_n \quad = \quad K_n + I_n \mathbb{E}_{\nu_n}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1\right], \tag{40}$$

which can be rearranged as

$$K_n \quad = \quad I_n\left(1 - \mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1\right]\right) = I_n f_n. \tag{41}$$

We now need to calculate the expectation value:

$$\mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1\right], \tag{42}$$

where r is the discount rate and where ν_n is the random time between consecutive investments. This expectation can be rewritten as

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1\right] = \\ &= s^{(n)}_{\text{high}} \frac{1}{(1+r)^n} + s^{(n)}_{\text{low}} \mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(n) = 0\right], \end{aligned} \tag{43}$$

where $s^{(n)}_{\text{high}}$ and $s^{(n)}_{\text{low}}$ are the probabilities of being in the high and low state respectively, in year n, conditioned on $\pi(0) = 1$. The distribution of ν_n, conditioned on $\pi(n) = 0$, is given by

$$P(\nu_n = k + n | \pi(n) = 0) = (1-p)^{k-1} p, \quad k \geq 1, \tag{44}$$

which is the probability of the profit π reaching the high state after k years for the first time after the current investment expires. The moment on the right-hand side in (43) is then found through

$$\mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(n) = 0\right] \quad = \quad \frac{1}{(1+r)^n} \sum_{k=1}^{\infty} \frac{1}{(1+r)^k} (1-p)^{k-1} p = \frac{1}{(1+r)^n} \frac{p}{r+p}.$$

Given an initial probability distribution, $s^{(0)} = (1, 0)^T$, (we start in the high state) the distribution in year k i given by $s^{(n)} = M^n s^{(0)}$. Assuming that $p \neq 1/2$, the transition matrix M is non-singular with eigenvectors $v_1 = (1, 1)^T$ and $v_2 = (1, -1)^T$ with corresponding eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - 2p$.With $s^{(0)} = \frac{1}{2}(v_1 + v_2)$ the distribution $s^{(n)}$ is then given by

$$s^{(n)} = \frac{1}{2} M^n (v_1 + v_2) = \frac{1}{2}(\lambda_1^n v_1 + \lambda_2^n v_2) = \frac{1}{2}\begin{pmatrix} 1 + (1-2p)^n \\ 1 - (1-2p)^n \end{pmatrix}$$

The expectation value in (43) can now be written as

$$\begin{aligned} \mathbb{E}\left[\frac{1}{(1+r)^{\nu_n}} \middle| \pi(0) = 1\right] \quad &= \quad \frac{1}{2}(1 + (1-2p)^n) \frac{1}{(1+r)^n} + \frac{1}{2}(1 - (1-2p)^n) \frac{1}{(1+r)^n} \frac{p}{r+p} = \\ &= \quad \frac{1}{(1+r)^n} \frac{1}{2}\left[1 + \frac{p}{r+p} + \frac{r}{r+p}(1-2p)^n\right]. \end{aligned} \tag{45}$$

The function f_n in (41) is therefore

$$f_n = 1 - \frac{1}{(1+r)^n}\frac{1}{2}\left[1 + \frac{p}{r+p} + \frac{r}{r+p}\left(1-2p\right)^n\right]$$

References

[1] About us. `http://www.gocpc.com/about.html`. Community Power Corporation.

[2] Applications of hpm. `http://www.hyperionpowergeneration.com/applications/`. Hyperion Power Generation.

[3] Cloromat fact sheet. `http://www.gewater.com/pdf/Fact20Sheets_Cust/Americas/English/_FS1298EN.pdf`. GE Power & Water: Water Process Technologies.

[4] Distributed generation: The use of biomass to create electricity. `http://tef.tulane.edu/pdfs/2011/jim-neumeier.pdf`. Innovative Energy Inc.

[5] Miox at-a-glance. `http://www.miox.com/about-miox/At-A-Glance.aspx`. Miox Web Site.

[6] Proven biomass gasification solutions by aesi. `http://www.aesintl.net/solutions/proven-biomass-gasification-solutions-by-aesi`. AESI.

[7] Remote controlled chlorine production. `http://www.akzonobel.com/ic/products/remote_controlled_chlorine_production/`. AkzoNobel Industrial Chemicals.

[8] U.s. nuclear operating plant basic information.

[9] Why modular biopower? `http://www.gocpc.com/biopower.html`. Community Power Corporation.

[10] *Distributed Generation in Liberalised Electricity Markets.* IEA, 2002.

[11] Renewable & alternative fuels net summer capacity. `http://www.eia.gov/cneaf/alternate/page/renew_energy_consump/table4.html`, August 2010. U.S. Department of Energy.

[12] Thinking small. `http://www.economist.com/node/17647651`, December 2010. The Economist Technology Quarterly.

[13] `http://minerals.usgs.gov/minerals/pubs/mcs/2011/mcs2011.pdf`, 2011. Mineral Commodities Summaries 2011.

[14] Deep sea fission. `http://www.world-nuclear-news.org/NN_Deep_sea_fission_2001111.html`, January 2011. World Nuclear News.

[15] Ict price basket. `http://www.itu.int/ITU-D/ict/ipb/`, May 2011. International Telecommunication Union.

[16] *Social Ramifications of Autonomous Urban Land Vehicles.* IEEE International Symposium on Technology and Society, May 2011, Chicago, 2011. Tyler C. Folsom QUEST Integrated, Inc.

[17] Uhdenora to supply uhde with skid-mounted chlor-alkali electrolysis modules for leuna- harze. `http://www.uhdenora.com/details.asp?id_news=73`, March 2011. UHDENORA web site.

[18] Sally Adee and Erico Guizzo. Nuclear redux. *IEEE Spectrum*, 47:25–32, 2010.

[19] Linda Argote and Dennis Epple. Learning Curves in Manufacturing. *Science*, 247(4945):920–924, 1990.

[20] Kenneth J. Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173, 1962.

[21] W. Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):pp. 116–131, 1989.

[22] Drew Bellamy and Luka Pravica. Assessing the impact of driverless haul trucks in australian surface mining. *Resources Policy*, 36(2):149 – 158, 2011.

[23] R. Byron Bird, Warren E. Stewart, and Edwin N. Lightfoot. *Transport Phenomena.* John Wiley & Sons, Inc., 2002.

[24] Walt Bogdanich and Christopher Drew. Deadly leak underscores concerns about rail safety. `http://www.nytimes.com/2005/01/09/national/09rail.html`, January 9 2005. The New York Times.

[25] E. Bozorgebrahimi, R. A. Hall, and G. H. Blackwell. Sizing equipment for open pit mining ï£¡ a review of critical parameters. *Mining Technology (Trans. Inst. Min. Metall.)*, 112:A171, 2003.

[26] Bernard L. Cohen. *The Nuclear Energy Option.* Plenum Press, 1990.

[27] Doug Cure. Pharmachem invests in green energy in nc facility with aesi system. `http://www.aesintl.net/blog/aesi-news/pharmachem-invests-in-green-energy-at-nc-facility-with-aesi-system`, October 10 2011. AESI Biomass Energy Blog.

[28] Paul A. David. Clio and the economics of qwerty. *The American Economic Review*, 75(2):pp. 332–337, 1985.

[29] EIA. Form 860, annual electric generator report 2008.

[30] Menachem Elimelech and William A. Phillip. The future of seawater desalination: Energy, technology, and the environment. *Science*, 333(6043):712–717, 2011.

[31] Jan Willem Erisman, Mark A. Sutton, James Galloway, Zbigniew Klimont, and Wilfried Winiwarter. How a century of ammonia synthesis changed the world. *Nature Geoscience*, 1:636–639, 2008.

[32] J. P. Euzen, P. Trambouze, and J. P. Wauquier. *Scale-up Methodology for Chemical Processes.* Gulf Publishing Company, 1993.

[33] B. C. C. van der Zwaan F. Ferioli. Learning in times of change: A dynamic explanation for technological progress. *Environmental Science & Technology*, 43:4002–4008, 2009.

[34] FERTECON. `http://www.fertecon.com/latest_market_reports/FERTECON_latest_Ammonia_market_report.pdf`. FERTECON AMMONIA REPORT, A Weekly Review of the International Ammonia Market, 8 December 2011.

[35] C. Fritzmann, J. Lwenberg, T. Wintgens, and T. Melin. State-of-the-art of reverse osmosis desalination. *Desalination*, 216(1-3):1 – 76, 2007.

[36] Lauren F. Greenlee, Desmond F. Lawler, Benny D. Freeman, Benoit Marrot, and Philippe Moulin. Reverse osmosis desalination: Water sources, technology, and today's challenges. *Water Research*, 43(9):2317 – 2348, 2009.

[37] John Haldi and David Whitcomb. Economies of scale in industrial plants. *Journal of Political Economy*, 75(4):pp. 373–385, 1967.

[38] John J Hisnanick and Kern O Kymn. Modeling economies of scale: the case of us electric power companies. *Energy Economics*, 21(3):225 – 237, 1999.

[39] Thomas Hoff. Integrating renewable energy technologies in the electric supply industry: A risk management approach. Technical report, NREL, 1997.

[40] Kenneth K. Humphreys and Sidney Katell. *Basic cost engineering.* New York : M. Dekker, 1981.

[41] Rumy Husan. The continuing importance of economies of scale in the automotive industry. *European Business Review*, 97(1):38–42, 1997.

[42] Paul J. and Bartos. Is mining a high-tech industry?: Investigations into innovation and productivity advance. *Resources Policy*, 32(4):149 – 158, 2007.

[43] Michael W. Jack. Scaling laws and technology development strategies for biorefineries and bioenergy plants. *Bioresource Technology*, 100(24):6324 – 6330, 2009.

[44] B. M. Jenkins. A comment on the optimal sizing of a biomass utilization facility under constant and variable cost scaling. *Biomass and Bioenergy*, 13(1-2):1 – 9, 1997.

[45] J.R Jennings and (editor). *Catalytic ammonia conversion.* Plenum Press, New York, 1991.

[46] A.C. Kadak, R.G. Ballinger, T. Alvey, C.W. Kang, P. Owen, A. Smith, M. Wright, and X. Ya. Nuclear power plant design project: Phase 1 review of options and selection of technology of choice. Technical report, MIT, 1998.

[47] Christian Kaufmann. State of the internet report ï£¡ q2 2011. Technical report, Akamai Technologies, Inc., 2011. `http://www.denog.de/meetings/denog3/pdf/10-Kaufmann-State_of_The_Internet.pdf`.

[48] W.G. Koellner, G.M. Brown, J. Rodriguez, J. Pontt, P. Cortes, and H. Miranda. Recent advances in mining haul trucks. *Industrial Electronics, IEEE Transactions on*, 51(2):321 – 329, april 2004.

[49] V. Kuznetsov. Innovative small and medium sized reactors: Design features, safety approaches and r&d trend. Technical report, IAEA, 2004.

[50] André Langevin, Pontien Mbaraga, and James F. Campbell. Continuous approximation models in freight distribution: An overview. *Transportation Research Part B: Methodological*, 30(3):163 – 188, 1996.

[51] James Larminie and Andrew Dick. *Fuel Cell Systems Explained.* John Wiley & Sons Ltd, 2003.

[52] Richard Levin. Technical change and optimal scale: Some evidence and implications. *Southern Economic Journal*, 44:208–211, 1977.

[53] Charles E. Levy. Lessons learned from raser technologies' 'revolutionary' project. `http://energy.aol.com/2011/10/20/lessons-learned-from-raser-technologies-revolutionary-project/`, October 20, 2011. AOL Energy.

[54] Marvin Lieberman. Market growth, economies of scale, and plant size in the chemical processing industries. *The Journal of Industrial Economics*, 36:175–191, 1987.

[55] Amory B. Lovins, E. Kyle Datta, Thomas Feiler, Karl R. Rabago, Joel N. Swisher, Andre Lehmann, and Ken Wicker. *Small is Profitable.* Rocky Mountain Institute, 2002.

[56] Timothy J. MacFall and Gregory M. Engleston. Rigrodsky & long, p.a. and the egleston law firm file shareholder class action lawsuit on behalf of raser technologies, inc. shareholders. `http://eon.businesswire.com/news/eon/20111129006954/en/Raser-Technologies/RZTIQ.PK/securities-fraud`, November 29, 2011. Enhanced Online News.

[57] Aamer Madhani. Profits elude geothermal companies. `http://www.usatoday.com/news/washington/story/2011-10-05/profits-elude-clean-energy-companies/50674030/1`, October 6, 2011. USA Today.

[58] Alan McDonald and Leo Schrattenholzer. Learning rates for energy technologies. *Energy Policy*, 29(4):255 – 261, 2001.

[59] Jan A. Van Mieghem. *Operations Strategy: Principles and Practice.* Dynamic Ideas, Llc, 2008.

[60] William D. Nordhaus. The progress of computing. March 2002.

[61] Steven Oberbeck. Raser says it's victim of $100 million 'joke'. `http://www.sltrib.com/csp/cms/sites/sltrib/pages/printerfriendly.csp?id=52568877`, September 27, 2011. The Salt Lake Tribune.

[62] National Academy of Engineering. `http://www.engineeringchallenges.com/`. Grand Challenges of Engineering.

[63] G. Pepermans, J. Driesen, D. Haeseldonckx, R. Belmans, and W. D'haeseleer. Distributed generation: Definition, benefits and issues. *Energy Policy*, 33:787–798, 2005.

[64] D. J. Petersen, Tom LaTourrette, and James T. Bartis. `http://www.rand.org/pubs/monograph_reports/MR1324`, 2001. New Forces at Work in Mining: Industry Views of Critical Technologies.

[65] Daryl Robertson. The powerplant and industrial fuel use act of 1978: Fuel replacement. *Harv. Envtl. L. Rev*, 214:217–235, 1979.

[66] Bruno Sauvet-Goichon. Ashkelon desalination plant – a successful challenge. *Desalination*, 203(1-3):75 – 81, 2007. EuroMed 2006 - Conference on Desalination Strategies in South Mediterranean Countries.

[67] Rebecca Smith. Small reactors generate big hopes. `http://online.wsj.com/article/SB10001424052748703444804575071402124482176.html`, February 18 2010. Wall Street Journal.

[68] R. M. Solow, J. Tobin, C. C. von WeizsÃd'cker, and M. Yaari. Neoclassical growth with fixed factor proportions. *The Review of Economic Studies*, 33(2):79–115, 1966.

[69] G. Stanek, D. Langer, B. Mu? andller Bessler, and B. Huhnke. Junior 3: A test platform for advanced driver assistance systems. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 143 –149, june 2010.

[70] Galina Stolyarova. Nuclear power vessel launched. `http://www.times.spb.ru/index.php?action_id=2&story_id=31865`, July 6 2010. The St. Petersburg Times.

[71] Erich Strohmaier, Jack J Dongarra, Hans W Meuer, and Horst D Simon. The marketplace of high-performance computing. *Parallel Computing*, 25(13ï£¡14):1517 – 1544, 1999.

[72] Jeffrey Tomich. Kirkwood utility is branching out with biomas. `http://www.stltoday.com/business/local/article_71d52909-707b-5d5a-9a34-65423badde96.html`, December 10 2010. St. Louis Today.

[73] Haldor Topsoe. Press release: Topsoe contracts the world's largest ammonia plant. Viewed Dec 2010 at http://www.topsoe.com/news/News/2009/020709.aspx.

[74] M.A. Tribe and R.L.W. Alpine. Scale economies and the Ş0.6 ruleŤ. *Engineering Costs and Production Economics*, 10(1):271 – 278, 1986.

[75] Haruki Tsuchiya and Osamu Kobayashi. Mass production cost of pem fuel cell by learning curve. *International Journal of Hydrogen Energy*, 29(10):985 – 990, 2004. Fuel Cells.

[76] Matthew L. Wald. Administration to push for small modular reactors. `http://www.nytimes.com/2011/02/13/science/earth/13nuke.html?pagewanted=all`, February 12 2011. New York Times.

[77] Spectra Watermakers. `http://www.spectrawatermakers.com/`.

[78] C.M. White, R.R. Steeper, and A.E. Lutz. The hydrogen-fueled internal combustion engine: a technical review. *International Journal of Hydrogen Energy*, 31(10):1292 – 1305, 2006.

[79] Michelle K. Wittholz, Brian K. O'Neill, Chris B. Colby, and David Lewis. Estimating the cost of desalination plants using a cost database. *Desalination*, 229(1-3):10 – 20, 2008.

[80] T.P Wright. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3:122 – 128, 1936.

[81] Marcus Yam. Air force still using ps3 as 33rd largest computer. `http://www.tomshardware.com/news/usaf-supercomputer-ps3-playstation-cell,11727.html`, Deceber 6, 2010. Tom's Hardware.