The Theory–Practice–Entrepreneurship Gap

My colleague John Doyle once said to me, "The gap between theory and practice is a lot bigger in practice than in theory." In this column, I will share my journey across this gap through entrepreneurship, especially the mistakes I have made, in response to Thomas Parisini's invitation to share personal experiences on how control theory impacts applications. I do not conduct research to create start-ups. However, I'm excited when some of our research has the potential to make an impact in practice. I'm not interested in initiating just any start-up but, rather, a high-tech start-up that translates our own research into the marketplace. There are many kinds of entrepreneurial models, and my experience is limited to a particular type of entrepreneurship, one with an extra risk. Unfortunately, for an academic like myself who is interested in research impact (not start-ups per se), this risk is almost unavoidable and must be managed. Before I elaborate, let me first clarify my view on practice, research, and theory research.

Guest Author Steven Low

Steven Low is the Frank J. Gilloon Professor in the Department of Computing and Mathematical Sciences and the Department of Electrical Engineering at the California Institute of Technology and an honorary professor at the University of Melbourne. Before that, he was with AT&T Bell Laboratories, Murray Hill, New Jersey, and the University of Melbourne, Australia. He has been a corecipient of IEEE best paper awards; an awardee of the IEEE INFOCOM Achievement Award (2021) and the Association for Computing Machinery (ACM) SIGMETRICS Test of Time Award (2021); and is a Fellow of IEEE (2008), ACM (2020), and the Chinese Society for Electrical Engineering (2020). He was well known for pioneering a mathematical theory of Internet congestion control and semidefinite relaxation of optimal power flow problems in smart grids. His research on networks has been accelerating more than 1 TB of Internet traffic every second since 2014. His research on smart grids is providing large-scale, cost-effective electric vehicle charging to workplaces. He received the B.S. degree from Cornell University and the Ph.D. degree from the University of California, Berkeley, both in electrical engineering.

Thomas Parisini

I respect scholarship and love research, like most academics do. Not everything useful, however, is, or needs to be, research. Not all good research has, or needs to have, practical use. I'm interested in their intersection. However, I also have a lot of admiration for people working in their complements. The primary goal of academia should be to train researchers and conduct good research, whether it has practical use or not. Why not invest only in good research that is useful in practice? Unfortunately, this strategy fails for two reasons. It is true that a lot of good research was driven by practical needs—Harold Black's invention of negative feedback in the 1920s and Claude Shannon's invention of information theory in the 1940s were motivated by the desire to enable long-distance communication through amplification and coding. It was no accident that both were invented at AT&T Bell Labs, the center of the communications industry of the last century.

However, there are also, plenty of examples where good research started purely as an intellectual pursuit but subsequently found widespread (and sometimes unexpected) applications. Quantum mechanics was developed more than 100 years ago to understand the physical world, but it now underlies numerous electronic devices. Some results in number theory developed centuries ago have led to cryptography that is now the bedrock of e-commerce security. It can be difficult to predict useful research even a few years out. As a funny example, I was a student representative from 1991 to 1992 at Berkeley for faculty hiring in the Electrical Engineering Department. A faculty candidate declared during a meeting with our student committee that "information theory is dead," only to see a massive revival a few years later as wireless communications swept through the world.

The second reason we cannot restrict research investment to only what we believe will have practical use

Digital Object Identifier 10.1109/MCS.2021.3107741
Date of current version: 12 November 2021

is that the research enterprise is like a tree. We cannot expect pretty flowers and juicy fruit without a strong root system and a healthy trunk, branches, and leaves. Good research, especially good theory research, sustained over time is one of the best ways to nurture intellectual fearlessness, elegant taste, and a focus on fundamentals. Such a culture is critical for training researchers who not only are capable and innovative but also possess a habit of cutting through unimportant details to the core of a problem. These qualities are portable beyond theory and beyond research, and they are necessary for tackling many complex problems of our time.

The gap between a research result and a viable business is summarized by what a patent attorney told me when we started our first company, "An idea is not a technology, a technology is not a product, a product is not a company, a company is not a business." Connecting one end to the other is what entrepreneurship is about. The right way to start a company is to identify a big existing or emerging pain point, identify a weak or missing link in its solution, build a business model, and then develop the necessary technology. In contrast, the type of start-up that I'm interested in necessarily begins with a technology and looks for a market need. This is hard, especially if the founding team consists only of people from academia and knows little about start-ups, markets, and management. It will be tremendously helpful if the founding team includes a seasoned entrepreneur who is passionate about the technology.

Research and start-ups are extreme points that are complementary yet similar. Both are intellectual, though to different degrees. Both are exciting because you work with a highly motivated and talented team. You have to develop a simple and compelling vision and convince people about it (students/employees, communities/customers, program managers/investors). In both cases, you must push forward despite uncertainty and without sufficient resources. At a start-up, however, the degree of uncertainty under which numerous decisions (big and small) must be made is larger, the lack of resources more severe, the speed at which decisions and adjustments are carried out higher, and their consequences often stronger. I have a lot more respect for good entrepreneurs and managers after the experience of my first start-up. I have come to realize that what they do is not only useful, but also difficult. Operating a good business takes more than developing a technology and productizing it. Managing a team of people, especially people with diverse talents and interests, to all row in the same direction is critical for a start-up. However, it is harder than inspiring brilliant young students to pursue whatever research that excites them. At Caltech, we have the rare privilege of not having to manage.

What can an academic hope to gain from starting a company? Three things, with decreasing probability: 1) to understand and practice entrepreneurship, 2) to see your research making an impact in practice, and 3) to be rewarded financially. I had no idea what a start-up was—that was an important reason for doing my first. No matter how much you read about start-ups, creating one is the only true way to learn. I learned so much about the theory–practice gap and what it takes to create a high-tech start-up. It boils down to market, team, technology, and funding, in that order. Like research, it takes a lot of passion and hard work.

RESEARCH

Our first start-up commercialized our research on Internet congestion control, a topic that I started to work on in the late 1990s at the University of Melbourne, Australia. After joining Caltech in 2000, I worked closely with John Doyle, Fernando Paganini (then next door at the University of California, Los Angeles), and others to develop a theory of Internet congestion control based on feedback control and convex optimization. It was an exciting time! The most important value of theory, more than yielding any particular algorithm, is that it brings clarity and structure to a problem. It provides a framework to understand issues, clarify ideas, and suggest directions, often leading to a more efficient, more robust, and simpler implementation.

The congestion-control mechanism in the TCP has been responsible for maintaining stability as the Internet scaled up in size, speed, traffic volume, coverage, and complexity by many orders of magnitude over the last three decades. The Internet, called the *Arpanet* at the time, was born in 1969 with four nodes. The TCP was published by Vinton Cert and Robert Kahn in 1974, split into the TCP/Internet Protocol (IP) in 1978, and deployed as a standard on the Arpanet by 1983.

An Internet congestion collapse was detected in October 1986 on a 32-Kb/s link between the University of California Berkeley campus and the Lawrence Berkeley National Laboratory 400 yards away, during which the throughput dropped by a factor of almost 1000 to 40 b/s. Two years later, Van Jacobson implemented and published the congestion-control algorithm in the Tahoe version of TCP [1], based on the additive-increase, multiplicative-decrease idea of Jain et al. [2].

Before Tahoe, there were mechanisms in the TCP to prevent senders from overwhelming receivers. However, there was no effective mechanism to prevent the senders from overwhelming the network. This was not an issue because there were few hosts until the mid-1980s. By November 1986, the number of hosts was estimated to have grown to 5089. However, most of the backbone links have remained at 50–56 b/s since the beginning of the Arpanet. Van Jacobson's scheme adapts sending rates to the congestion level in the network, thus preventing the senders from overwhelming the network. It worked very well over networks with relatively low transmission capacity, small delay, and few random packet losses. This was mostly the case through the 1990s. However, as the network speed underwent rapid upgrades (Figure 1), as the Internet exploded onto the global scene beyond research and education, and as mobile services proliferated on the Internet, the strain on the original design started to show.

This prompted a flurry of research on TCP congestion control starting in the early 1990s. A mathematical understanding of network congestion control started with Frank Kelly's 1997 article on network utility maximization [3]. I was visiting Albert Greenberg at Bell Labs, Murray Hill, in December 1997 when he told me about Kelly's idea. I worked out a simple design based on duality theory during my visit and gave a talk at Berkeley in February 1998 on my way back to Melbourne. The idea was to view TCP senders and network routers as carrying out a distributed computation over the Internet in real time to solve the dual problem of Kelly's utility maximization. This basic design and an IP implementation by my student David Lapsley were described in a pair of conference papers in 1998 [4], [5]. The implementation was a proof of concept but not practical as it requires the network to feed back a multibit congestion signal to a TCP sender, whereas Internet standards allow feedback of at most a single bit.

This motivated our invention of random early marking in 1999, which allowed a TCP sender to estimate a congestion signal probabilistically from single-bit feedback [6]. I felt confident about this approach after I visited Larry Peterson at Princeton in 1999 to finish reverse-engineering TCP Vegas using the duality model (unfortunately, our conference submission was rejected). Before the rise of the Internet in the 1990s, telephony dominated communications, and the main theoretical framework was queueing theory (dating back 100 years to the work of the Danish mathematician Agner Erlang). However, it is difficult to incorporate into queueing theory the distributed real-time feedback control that is the hallmark of the Internet. By 2000, it was clear that a new framework based on control theory and convex optimization was needed to think about Internet congestion control. An intensive effort ensued in the control and networking communities that lasted a decade to reverse-engineer existing algorithms, understand the structural properties of large networks under end-to-end control, systematically design new algorithms based on analytical insights, and deploy some of these innovations in the field.

By early 2002, even though the theory still had many unresolved questions, I felt there was sufficient understanding that we could test various ideas in a real network. Most of my lab continued to work on theory. However, I also started a small team that initially consisted of a student, David Wei, and a postdoc, Cheng Jin, to implement our new FAST TCP algorithm, which has the same equilibrium as TCP Vegas. FAST TCP was much more efficient than the standard TCP algorithm over high-speed, long-distance networks and wireless links. We worked with Harvey Newman and his collaborators at CERN, Internet2, SLAC, and other research labs around the world to test our implementation. Our demonstration at the SuperComputing Conference in November 2002 was a major success and helped change how the Internet Land Speed Record was conducted. For example, subsequent to our demo, the TCP transfer speed had to sustain for at least an hour, instead of seconds to minutes, because of the fragility of earlier attempts to break data-transfer records. We rapidly expanded our research program, pushing on theory, implementation and testbed, expanding our collaborations, and exploring opportunities for deployment including through Internet Engineering Task Force standardization.

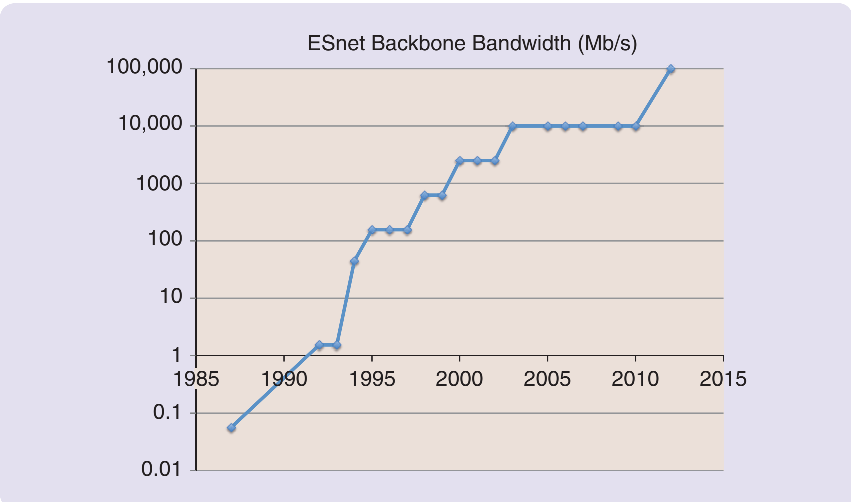

FIGURE 1 The highest link speed of the U.S. Department of Energy's ESnet from 1987 (56 Kb/s) to 2012 (100 GB/s). ESnet: Energy Sciences Network.

PRACTICE

When we finally started a company in February 2006 to deploy our research on our own terms, I thought, mistakenly, that our research prototype was ready for prime time. Our technology was a piece of TCP kernel software. We deployed it as a network appliance so that it could be inserted into a corporation's data center without modifying its operating system (Figure 2). After failing miserably in the data centers of our first three beta customers, I realized that we did not have a product—or even a technology—that was ready for productization. We suspended our beta testing program and recruited a student, George Lee, to build a serious testbed at the start-up and then around the world. Our appliance touched all kinds of hardware and software that existed in a corporation's data center. In theory, standard protocols should ensure seamless integration. In practice, a transport-layer product must operate defensively to survive all of the corner cases it will encounter. Kernel programming is notoriously difficult to debug. To develop such a robust product, we needed more than just superb in-house development and testing; we also had to throw it into as many customer networks as we could recruit. This created a chicken-and-egg problem: without a robust product, it was hard to convince corporations to allow us into their networks, which made it hard to build a robust product.

The difficulty extended far beyond product robustness. It took a while, but we finally developed an appliance that could routinely accelerate TCP connections by five to 30 times between our customers' data centers and that was extremely resilient to packet loss (Figure 3). Since the TCP carried more than 90% of the Internet traffic, it seemed obvious to me that every company would need our product. I did not appreciate that, unlike in research, where pushing the boundary is an end in itself, a commercial company looks for value. A higher performance is of no value if it does not translate into lower costs or higher revenues for the company. This required us to identify a sector with a pain point for which our technology is a must-have, understand how companies in this sector operate, quantify the value our product could provide them, and evaluate how our own start-up could make a profit. In other words, we needed to develop a business model and a go-to-market strategy. This is nontrivial for seasoned entrepreneurs; as a novice, I did not even realize that it was hard.

We quickly raised significant angel funding and sold our appliances to several Fortune 100 companies at a good profit. This early success turned out to be a curse in disguise. It was due more to the excitement about our technology than a strong business model. For example, we went after the entertainment industry early on because of their need to transfer large video files. That was, however, the time when Netflix's business was exclusively shipping DVDs. Streaming video on the Internet would have been a perfect market for our technology, but it was not yet a large business around 2006. While being ahead of its time is great for research, it can be devastating for a start-up. To create a market and the necessary infrastructure would require a company of scale and complexity far beyond our ability.

In retrospect, we should have worked to better understand markets, figure out a suitable business model, and adapt our product accordingly. Instead, we pursued disparate use cases for our existing product without a clear strategy. One of our early customers was a movie department of a major entertainment studio in Los Angeles that wanted to replace DVDs with our appliance for transporting movies to South America. Our appliance needed to be deployed in the studio's data center, which was managed by the studio's IT department. It was difficult to convince the IT department to manage yet another piece of equipment when the addition provided no benefit, but only risk, to the IT department, especially if the purchase had to come out of its own budget. It was harder to convince the movie department to entrust its business operation to an unknown start-up, especially if the alternative solution is nice to have,

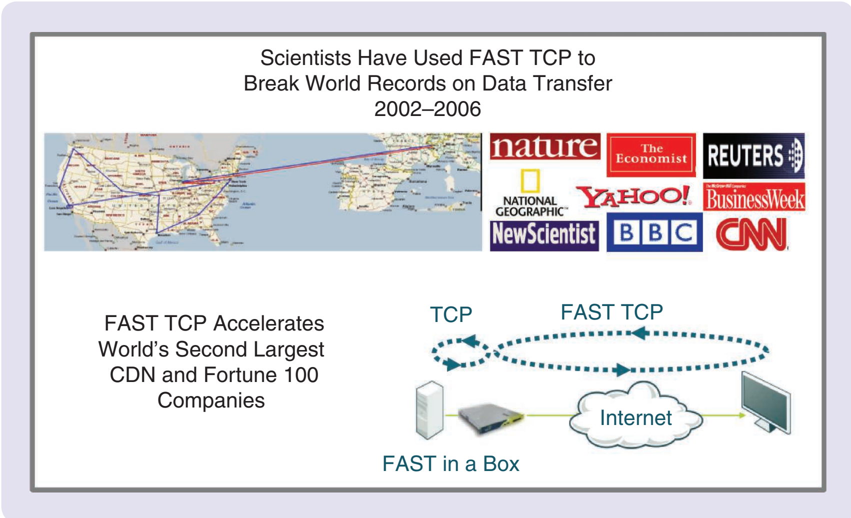

FIGURE 2 FAST TCP appliance and the start-up's pitch. CDN: content distribution network.

FIGURE 3 When the Sprint network experienced heavy packet losses in June 2007, FAST TCP increased throughput between San Francisco and New York by 120 times.

For a horizontal technology, like ours, that was applicable to many use cases, it can be particularly hard not to chase revenue opportunities that seem to pop up in all directions.

not must have. We had good revenue in 2007, but this early success did not help us figure out a strong business model and it was too complex to scale. My inexperience, however, mistook these early sales as a validation of our market, our company, and our product, and we decided to scale our operation. I thought my job at the start-up was accomplished. We hired a business team, moved to a bigger office, and I prepared to return to my research at Caltech. Then the financial crisis of 2008 hit.

Having been through a complete lifecycle of technology development, from a mathematical idea to real-world deployment, I restarted my research from scratch in power systems after I returned to Caltech. I still consulted for the start-up and witnessed its navigation through the recession. The rise of cloud computing around 2009 offered an excellent opportunity for our technology, but it would require a different business model. The DNA of the company and the recession made that transition difficult. One of the most common challenges that kills a start-up is a lack of focus. For a horizontal technology, like ours, that was applicable to many use cases, it can be particularly hard not to chase revenue opportunities that seem to pop up in all directions. These opportunities may help financially in the short term, but if they do not amount to an overall go-to-market strategy, they drain precious resources, the most critical of which is time, and delay the urgent need to nail a strong business model.

Focus requires not just discipline but, more importantly, the clarity of a business model. While the team must be nimble tactically in adapting to uncertainties, it must be precise and steady strategically. In the end, we decided that we created a valuable technology the market needed, and it would be better to let others scale its deployment. Our start-up was acquired in 2012. The technology was integrated in 2014 into the world's largest content distribution network and accelerated more than 1 TB of Internet traffic every second.

It was an exhilarating and exhausting journey.

REFERENCES

[1] V. Jacobson, "Congestion avoidance and control," in *Proc. SIGCOMM'88*, ACM, Aug. 1988, pp. 314–329. doi: 10.1145/52325.52356.

[2] R. Jain, K. K. Ramakrishnan, and D.-M. Chiu, "Congestion avoidance in computer networks with a connectionless network layer," Digital Equipment Corporation, Littleton, MA, Tech. Rep. DEC-TR-506, 1988.

[3] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, 1997. doi: 10.1002/ett.4460080106.

[4] D. Lapsley and S. H. Low, "An optimization approach to ABR control," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 1998, pp. 1500–1504. doi: 10.1109/ICC.1998.683074.

[5] D. Lapsley and S. H. Low, "An IP implementation of optimization flow control," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Nov. 1998, pp. 3023–3028.

[6] D. Lapsley and S. H. Low, "Random early marking for internet congestion control," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 1999, pp. 1747–1752. doi: 10.1109/GLOCOM.1999.832461.