Technology Center: **Communication for Science Center**
USICA Focus Area: Advanced communications technology and immersive technology
Author: Michelle Effros

Abstract: By bringing together scientists from fields where information needs present a critical bottleneck with experts from disciplines focused on efficient and reliable information storage and transmission, the proposed Communication for Science Center seeks to unleash advances in both communications and science that neither community is likely to develop on its current path of largely independent investigation. These advances, in turn, are expected to shepherd future technological and economic development in line with the objectives of the United States Innovation and Competition Act (USICA).

Background: For over a century, communication technology has been a key driver for the country's technological and economic advancement. Not only do communication technologies lie at the heart of our communication and computing industries, they also play a central enabling role for almost every aspect of our personal, economic, and societal advancement. Healthcare systems, power distribution networks, supply chains, transportation networks, educational institutions, governmental bodies, social networks, industrial plants, and markets, to name just a few, all not only rely so critically on current communication technology that they are brought down by communication network failures but also rely on future advancements in those technologies to unlock much of their own future progress.

Many scientific endeavors at Caltech fall into a similar pattern. Whether we are collecting data from the Mars rover, modeling the earth's climate, distributing data from LIGO, searching for variable radio emissions from astronomical objects like stars, brown dwarfs, and planets, sequencing the regulatory genomes of new species, or tracing the circuitry of the brain, scientists at Caltech increasingly find themselves in scenarios where the ability to gather the data they need, deliver and store it efficiently and reliably, and employ that data for scientific investigation represent critical bottlenecks in achieving our scientific objectives.

Proposal: Caltech's position at the forefronts of communications and a wide array of data-intensive science disciplines and its relationship with JPL place us in a unique position to pursue the unexplored boundaries of these disciplines. We illustrate this boundary through a few questions at this boundary.

- **What data should be gathered?** The field of communications was developed under the assumption that the goal of every communication system is to efficiently represent information gathered at some source and reliably deliver that information across time (for information storage) or space (for information delivery) to a receiver interested in its reconstruction. This view of communication is explicitly agnostic to the information itself; from the perspective of the traditional communication system designer, the data has been gathered, and the task at hand is reliable and efficient transmission or storage. In the communication scenarios that arise in scientific endeavors like remote sensing, neuroscience, astronomy, and seismology, the story is considerably more complex. Here, the question of which data to gather is increasingly intertwined with the communication task. The goal then is to use what we know now to steer what data we gather in the future. Such questions can arise in many ways and on many time scales, and they can inspire new communication strategies. For example, deciding where to build the next telescope, how to distribute testing resources for COVID-19 across time and space, and where to steer a sensor-gathering drone are all questions about what data to gather, and each changes the data representation and transmission problems to require more feedback for different objectives and to create new dependencies in the data that communication algorithms can exploit.

- **What information should be shared?** As scientific applications increasingly move to distributed collection strategies and technologies, the problem of using what we know now to steer what we collect in the future raises new communication challenges. In a distributed environment, no individual sensor may have sufficient information to determine what it or other devices should collect; only by working together and sharing information amongst themselves can data gathering devices make good decisions about future data collection. In applications like the Event Horizon Telescope and the Large Hadron Collider, where distributed data collection, massive data sets, and severely restrictive communication constraints conspire to slow information transmission to a time scale of months rather than milliseconds, decisions about what to communicate, when, and to whom, are irrevocably intertwined with the communication process itself. Determining how and which information to share in such a distributed system is an open problem in communications.

- **For what will the data be used?** While the traditional approach to science involves first gathering data and then analyzing the data received, this approach is often extremely inefficient from a communication perspective. What we really need is not measurements but answers to the questions for which those measurements are made. Recognizing this discrepancy between the data gathered and the function of the data that motivates its collection changes the communication problem, allowing distributed information processing that can vastly reduce bandwidths and improve communication times.

Solving each of these example problems requires new communication theory and technology and makes new science possible. Bringing together experts with the relevant scientific and communication expertise is expected to create a beneficial synergy that, to our knowledge, is not replicated elsewhere.