Ward Whitt

Ward Whitt (born in 1942) spent his early years in Bozeman, Montana, before attending Dartmouth (AB in mathematics) and Cornell (PhD in operations research, 1969). His doctoral thesis, *Weak Convergence Theorems for Queues in Heavy Traffic*, advised by D. Iglehart, set the direction for his future research. Whitt's career has three phases: In 1969–1977, he was at Yale; in 1977–2002, he was at AT&T, first in Bell Labs and then AT&T Labs; since 2002, he has been at Columbia in the IEOR Department.

Whitt's research has focused on stochastic queueing models and their application to communication, computer, manufacturing and service systems. In his 330 publications he has also contributed more broadly to stochastic modeling, e.g., to stochastic-process limits, stochastic comparisons, simulation methodology, numerical transform inversion, approximate dynamic programming and heavy-tailed probability distributions.

Whitt's 1970 papers with Iglehart were seminal contributions on the use of asymptotic methods to develop tractable approximations describing the performance of complex queueing systems. These papers show the power of functional limit theorems, the continuous mapping theorem and state space collapse to establish heavy-traffic limits for multi-server queues with non-exponential distributions and networks of such queues. Whitt's work culminated in his 2002 book, *Stochastic-Process Limits*, for which he was awarded the INFORMS Lanchester Prize. As stated in the preface, ''Since the converging processes are constructed by appropriately scaling time and space, they provide a macroscopic view of uncertainty. The stochastic-process limits strip away unessential details and reveal key features determining performance.''

Other notable stochastic-process limits are Whitt's 1981 paper with S. Halfin initiating the ''Halfin-Whitt'' many-server heavy-traffic regime, his 1985 paper establishing the asymptotic correctness of the Erlang fixed-point approximation for stochastic loss networks, complementing work by F. Kelly, his 1991 paper with P. Glynn establishing hydrodynamic limits for the output from many empty queues in series, and his 1994 paper with P. Glynn establishing the fundamental large deviations theory for non-Markovian queueing models, supporting the concept of effective bandwidths for communication networks, which Whitt studied further with A. Berger.

Whitt was strongly influenced by the practical engineering environment of Bell Labs. In the early 1980's that motivated him to develop the *Queueing Network Analyzer* (*QNA*) software package, implementing a decomposition approximation to describe the steady-state performance of a non-Markovian open network of queues. The arrival process and the general service distribution at each queue are each approximately characterized by two parameters, one describing the rate or mean and the other describing the variability. The arrival-process variability parameters are obtained by solving a system of equations, like the traffic rate equations for the arrival rates, based on a ''network calculus'' to represent the operations of superposition, splitting and flow through a queue. That tool motivated Whitt's research during that period, including studies of how to approximate a stationary point processes by a renewal process. Papers with K. Sriram and K. Fendick showed how to capture the performance impact of dependence among interarrival and service times, as occurs with superposition arrival processes.

The need for useful practical results also motivated Whitt to work with J. Abate and other Bell Labs colleagues G. Choudhury, D. Lucantoni and K. Leung to develop effective algorithms for numerically inverting transforms and apply them to a wide array of queueing models. These results refuted the conventional wisdom that many previously derived transforms were of little value. Their work received honorable mention for the 1997 INFORMS Lanchester Prize.

Whitt investigated many aspects of queueing systems. With W. A. Massey and others, he studied the performance of multi-server queues with time-varying arrival rates. That work is surveyed in a 2007 paper in this journal, co-authored with L. V. Green and P. J. Kolesar. He helped improve understanding of fundamental queueing principles: Little's law (with P. Glynn) and Poisson Arrivals See Time Averages (with B. Melamed).

Since joining Columbia in 2002, Whitt has developed models to analyze the performance of large-scale service systems, such as customer contact centers and healthcare systems, mostly in collaboration with students: on simulation-based staffing algorithms with R. Wallace, on ratio controls with I. Gurvich and O. Perry, on two-parameter processes with R. Talreja, G. Pang and Y. Liu, and on delay prediction with R. Ibrahim.

Whitt was recognized for his research contributions by several awards: the National Academy of Engineering in 1996, AT&T Fellow in 1997, INFORMS von Neumann Prize in 2001 and INFORMS Inaugural Fellow in 2002.

Ward Whitt