A First-Principles Framework for Networking Education

Jaber Daneshamooz, Sanjay Chandrasekaran, Arpit Gupta
University of California Santa Barbara

ABSTRACT

Generative AI trivializes the production tasks that protocol-survey networking courses assess. *Claim evaluation*, recognizing plausible-sounding wrong claims about unseen systems, becomes the surviving educational target. We evaluate a first-principles framework: four invariants (State, Time, Coordination, Interface), the Environment-Measurement-Belief decomposition inside State, and binding-constraint cascade reasoning, deployed as the spine of an upper-division networking course. A within-subject end-of-course survey, in which each student judges the course against their own prior course ($n = 40$ research-consenting students, 39 of whom took the protocol-survey prerequisite at the same institution), supports three findings. First, the framework produces measurably deeper understanding of system design rationale (Q1: $t = +11.48$ against neutral, $p < .001$, the largest effect size in the survey). Second, it supports analytical transfer to unseen systems (Q7: $t = +8.21$) and a specific claim-evaluation win: four respondents independently caught the same wrong claim about WiFi throughput, three of them naming the same correct mechanism. Third, it identifies the framework's current edge: its analytical mode markedly outpaces its generative mode (Q12: $t = +5.60$, the smallest effect size in the survey), motivating the next round of framework-design work.

1 INTRODUCTION

Content target outpaced assessment instruments. Advanced-networking courses today are organized as protocol surveys (chapter-per-topic, page-per-protocol) patterned on Kurose-Ross [13], which aims at principles but anchors its assessments in protocol description. Students who pass such a course can describe TCP [12] and CSMA/CA (WiFi's medium-access scheme) [2, 10] but cannot reason about a protocol they have not yet seen. When the next protocol arrives (QUIC [14], 802.11ax [11], or whatever follows), protocol-survey-trained students learn it from scratch. The gap was tolerable when course material was the bottleneck. It is not tolerable now.

The AI inversion. Large language models [3] produce correct-sounding TCP header walkthroughs and CSMA/CA backoff traces in seconds. Assignments and exams that test *production* of protocol descriptions are AI-trivializable. What AI does *not* reliably do is *claim evaluation*: recognizing when a confident-sounding claim about an unseen system is structurally wrong. Computing has become a generative discipline [6]; pedagogy must teach what produces claim evaluation if claim evaluation is what survives.

Structural questions, not specific answers. A student who has learned only protocol facts cannot evaluate a claim about a protocol they have not yet seen. A student who has learned the *structural questions* every networked system must answer can do so: they derive what an architecture must look like under a given constraint and reject claims that violate the derivation. The first-principles framework this paper evaluates [7, 8] encodes those questions in three operational components: four invariants (State, Time, Coordination, Interface), the Environment-Measurement-Belief decomposition inside State that diagnoses why systems fail, and the binding-constraint cascade that turns one environmental constraint into a derivation of the rest of the architecture (§2).

Contributions. Three contributions, anchored to an end-of-course survey ($n = 40$ research-consenting students; §3). **(1) Deeper understanding.** The framework outperforms protocol-survey instruction on the within-subject prior-course comparison: Q1 reaches $t = +11.48$ ($p < .001$), the largest effect size in the survey; 40% (16/40) of free-text responses corroborate by naming the prerequisite (§4.1). **(2) Transfer and claim evaluation.** The framework supports analytical transfer (Q7: $t = +8.21$; three transfer clusters at 28%, 28%, 18%) and enables a specific claim-evaluation win: four respondents independently caught the same wrong claim about WiFi throughput, three naming the same overhead mechanism (§4.2, §4.3). **(3) The framework's edge.** The analytical mode (decompose an instance) markedly outpaces the generative mode (predict from a constraint): Q12 carries the smallest effect size ($t = +5.60$), corroborated by 12% boundary-fuzziness reports; naming this asymmetry motivates the next round of framework-design work (§4.4).

2 THE FRAMEWORK

The framework has three operational components. The *four invariants* are the questions every networked system must answer. The *Environment-Measurement-Belief (E-M-B) decomposition* within the State invariant is the diagnostic for why systems fail. The *binding-constraint cascade* is the predictive tool that turns a single environmental constraint into

a derivation of the rest of the architecture. Each component is defined below as an operational rule (something a student does), with a worked example drawn from a system the course examines in depth.

2.1 Four Invariants

Every networked system answers four invariant questions; the questions stay constant, the answers differ. *State*: what does the system know about its environment, and where does the knowledge live? *Time*: what cadence drives feedback: event-triggered or timer-triggered, prescribed or inferred? *Coordination*: who decides among the parties, one entity, many, or a hierarchy? *Interface*: what is exposed across each boundary, and what is hidden? The four questions are complete in a testable sense: proposed additions reduce to joint answers along them (naming is an Interface-and-State choice, security a State-and-Coordination one), so the set is a falsifiable basis rather than a vocabulary. Different binding constraints cascade through these four answers and produce different architectures from the same template. *TCP* [12] inherits Interface from IP's best-effort contract (delivery attempted, nothing guaranteed), so Coordination is distributed, State is inferred from ACKs, and Time is event-driven on ACK arrivals. *Medium access (WiFi's Distributed Coordination Function, or DCF)* [2, 10] starts at Coordination because wireless transceivers cannot hear their own transmission; State stays local, Time is slotted with randomized backoff. *Adaptive-bitrate (ABR) video streaming* [9, 15] anchors at Time: the absence of a live deadline permits a 60–200 s client-side buffer that becomes the design's central element. *Routing* is locked at Coordination by Baran's 1964 survivability argument [1], which forbids a central point of control and yields distributed routing within an autonomous system (AS), coordinated across systems by the Border Gateway Protocol [16].

2.2 E-M-B Decomposition

E-M-B is the internal structure of the State invariant, where failures become observable. It turns *why systems fail* into a three-layer inspection. *Environment* is what is true. *Measurement* is what the system observes. *Belief* is what the system concludes. Failures are gaps between adjacent layers. Bufferbloat [5] is the standard example: environment (queue full), measurement (loss absent because the buffer absorbs packets), belief ("no congestion, send faster"). The gap is level-scoped: each flow and the router each hold an intact E-M-B chain, and the divergence surfaces only at the composed end-to-end level, where real latency outruns every endpoint's belief. The same decomposition transfers to any feedback-driven system that infers its state from a measurement, such as a ride-sharing application inferring a driver's location from GPS.

2.3 Binding-Constraint Cascade

The four invariants are the coordinate system of the design space; a binding constraint carves an admissible region within it, and each working system is a point in that region. Locking one invariant narrows the admissible answers for the others, which is the cascade, and one region can hold several valid designs at once. WiFi and cellular manage the same problem (a shared wireless channel) under different anchors: WiFi's unlicensed medium [10] locks Coordination to distributed, forcing State local and Time coarse; cellular's licensed spectrum locks Coordination to centralized, permitting State global and Time precise. The cascade is predictive: 802.11ax adopts centralized scheduling on an unlicensed band [11] because user density makes distributed coordination untenable, the cascade's prediction once Coordination becomes binding. This worked example was independently named by 11 of 40 respondents (28%) as the moment the framework "clicked" (§4.2).

2.4 Cascade as a Claim-Evaluation Check

The framework is applied in two directions. *Decomposition* takes a system or a proposed design apart into its invariant answers; *composition* derives those answers from a binding constraint. Claim evaluation is decomposition turned on a proposed claim; the evidence section (§4) calls decomposition the analytical mode and composition the generative mode. The framework supports claim evaluation because cascade-based prediction is a falsifiable check on any claim about an unseen system. A student runs the check in four steps: (1) identify the binding constraint the claim implicitly assumes, (2) compute what the cascade predicts under that constraint, (3) compare the claim against the prediction, (4) reject the claim if it violates the prediction. Consider "increasing WiFi's raw bit rate (its PHY rate) solves the throughput problem". At high bit rates, fixed-duration protocol overhead (the mandatory idle gaps and acknowledgment that bracket every frame: DIFS, SIFS, ACK; see [2]) dominates the per-frame cost; the cascade predicts throughput bounded by that overhead regardless of bit rate; the claim violates the prediction and is structurally wrong. §4.3 reports how many students independently ran exactly this check. Large language models [3] produce plausible-sounding claims that pass surface consistency checks. The cascade check asks a different question, whether the claim is consistent with the structural derivation from a known constraint, and that is the claim-evaluation skill the framework targets.

3 METHOD

Deployment and design. The framework was deployed as the spine of an upper-division advanced-networking course at UC Santa Barbara: 16 lectures, an open-access textbook

(8 chapters), a midterm, and a final project, all organized around the four invariants, E-M-B, and the cascade. The lecture format puts the question before the answer; the lecture notes are self-contained; assessments test application of the framework to NEW scenarios. The cohort gives a within-subject control: 39 of 40 research-consenting respondents took the protocol-survey prerequisite at the same institution, so the comparison items Q1 ("deeper understanding of WHY than my prior course") and Q7 ("better able to analyze a system I have NEVER seen") compare framework-based instruction against the respondent's own prior-course experience. Within-subject comparison holds the instructor, instrument, era, and student selection constant; it is not a randomized control, but it is the cleanest design available short of one. It does not control for course level, so we read Q1 and Q7 as suggestive rather than dispositive.

Instrument and analysis protocol. An anonymous end-of-course survey deployed in Week 10 collects 12 five-point Likert items (1 = strongly disagree, 5 = strongly agree) and 9 free-text items. The full text of the 12 Likert items is in Appendix A; this section refers to them by Q-number. Of 61 responses, 40 carry explicit research consent with no countermanding "course-improvement-only" selection; the 9 that selected both options (semantically contradictory under the form's multi-select) we treat as opt-out. Quantitative claims use $n = 40$. The analysis protocol is pre-stated: Likert items are tested with a one-sample t-test against the neutral midpoint of 3 (df = 39), and all 12 items clear $p < .001$, so the contribution rests on the *rank-ordering* of effect sizes (with n fixed across items, each item's t measures how far its mean sits above the neutral midpoint), not on pass/fail of any single item. Free-text items use a pre-stated $\geq 10\%$ ($\geq 4/40$) threshold for paper-relevant claims; single-mention findings are excluded regardless of vividness. Threshold and protocol were fixed before the contribution claims were drafted.

4 EVIDENCE

Each subsection runs the same move sequence: a quantitative anchor, a qualitative cluster meeting a pre-stated $\geq$ 10% threshold, and an explicit statement of what the data does and does not license. Every Likert item is significantly above the neutral midpoint at $p < .001$; the contribution lives in the rank-ordering of effect sizes, not in pass/fail of any one item.

4.1 Deeper Understanding

Anchor. Q1 ("deeper understanding of WHY than my prior course") has mean 4.33, $t = +11.48$ against neutral ($p < .001$): the largest effect size in the survey. Q2 (E-M-B helped me understand WHY systems fail) reaches $t = +10.44$ and Q3 (failure-driven generational framing) reaches $t = +9.09$. The three highest-effect-size items all concern *why* systems are designed as they are (Figure 1).

Corroboration. Forty percent of free-text responses (16/40) name the prerequisite when describing the difference. R33: "CS 176A mostly taught 'this is how a system/protocol works'; CS 176C definitely has a larger emphasis on 'why this system/protocol works the way it does.' " R21: in the prerequisite "we are told information"; in this course "I can look at systems I've never seen before and reason about it".

Takeaway. Holding the cohort constant, framework-based instruction produces measurably deeper understanding of design rationale than protocol-survey instruction. The largest effect size in the survey plus 40% qualitative corroboration (against the same students' own prior experience) rule out the common cohort, instructor, and instrument confounds.

4.2 Analytical Transfer

Anchor. Three convergent Likert items support transfer. Q7 (analyze a system I have NEVER seen) reaches $t = +8.21$, Q5 (same patterns across systems) reaches $t = +8.42$, and Q6 (midterm tested NEW scenarios) reaches $t = +8.27$. They target self-report, cross-system pattern recognition, and assessment-instrument validation respectively; convergence reduces the chance of single-item artifact.

Corroboration: three clusters. Three qualitative clusters meet the $\geq$ 10% threshold. *TCP/UDP decomposed through the framework* appears in 11/40 (28%); R13: "originally I would have just understood the TCP protocol as a procedure of steps but now I understand the reasoning." *WiFi/cellular centralization shift* appears in 11/40 (28%); R36: "Both WiFi and SDN [software-defined networking] started with distributed coordination and centralized due to facing the same pressure." *Bufferbloat [5] as E-M-B failure* appears in 7/40 (18%); R21: "a closed loop system of E-M-B measuring, thinking, and reacting."

Takeaway. The framework supports decomposing systems beyond those walked through in lecture, including the midterm's load-balancer scenario, a non-networking system analyzed structurally. This demonstrates transfer to unseen systems that the introduction framed as absent in protocol-survey-trained students.

4.3 Claim Evaluation

Anchor. Q10 (framework helps me evaluate claims about systems) reaches $t = +7.66$ and Q11 (framework vocabulary improves my AI-tool interactions) reaches $t = +5.78$; both positive at $p < .001$ with effect sizes smaller than the analytical-mode items. Students confirm the framework helps; they stop short of claiming routine adversarial mastery.

Corroboration: WiFi overhead-tax. The cleanest specific claim-evaluation case is recognition that "increasing WiFi's PHY rate solves the throughput problem" is plausible-sounding but wrong. Four respondents (10%) named it independently; three named the same correct mechanism. R36: "protocol overhead is still about 160 μs per frame, regardless of PHY rate." R30: "timing costs like DIFS, SIFS, ACKs, which I learned is known as the overhead tax." R39: "not just about speed, but a fixed overhead and too many devices contending for bandwidth."

Takeaway. Convergent diagnosis of the same wrong claim, by independent students, naming the same underlying mechanism, is evidence of *transferable* claim-evaluation capability, not topic-specific recall. Topic-specific recall produces paraphrases of a single lecture moment; convergent diagnosis of a claim *not stated in lecture*, through application of framework primitives stated in lecture, is the structural pattern that distinguishes claim evaluation from recall. The claim is correlational, not causal: the within-subject design controls the prior-course confound but does not establish that the framework *causes* the claim-evaluation capability.

4.4 The Framework's Edge

Anchor. Q12 (predict architecture from binding constraint) has $t = +5.60$, the smallest effect size in the survey, positive at $p < .001$. The contrast with the analytical-mode transfer item Q7 ($t = +8.21$) is sharp: decomposing a given system has approximately 1.5× the effect size of predicting an architecture from a constraint alone.

Corroboration. Five of forty respondents (12%) flag invariant boundaries as fuzzy when applied to protocols not decomposed in lecture. R25: "the framework felt a bit ambiguous and hard to define for new protocols." R33: "the boundaries between the three components of E-M-B" felt "rather murky." Invariant categories are operationally crisp for instance-decomposition; operationally fuzzy for constraint-driven prediction.

Takeaway. The framework's analytical mode is substantially more developed than its generative mode, a structural property of the framework as currently formulated, not of this offering. Closing the gap requires articulating predictive rules of the form "when constraint X is binding, invariant Y locks first and forces commitments Z" with the same operational crispness as the analytical categories. Naming the gap is the paper's third contribution.

5 DISCUSSION

Scope of claims. The survey supports three findings, the third an honest limit. The framework produces deeper understanding than protocol-survey instruction (Q1, $t = +11.48$, 40% corroboration). It supports analytical transfer (Q7, $t = +8.21$; three clusters at 28%/28%/18%) and specific claim-evaluation wins (Q10, Q11 significant; WiFi-overhead cluster). Its analytical mode outpaces its generative mode (Q12, $t = +5.60$, 12% boundary-fuzziness). The claims are correlational: the within-subject design controls the prior-course confound but does not establish causation, and it cannot separate the framework from course level, since any advanced second course may deepen understanding on its own. The framework-specific weight therefore rests on the mechanism-naming convergence (§4.3) and the analytical-generative asymmetry (§4.4), which a generic advanced course would not produce. Generalization beyond this cohort, instructor, and institution requires replication. The headline, "framework-trained students demonstrate framework-mediated claim evaluation," describes what we observe, not routine adversarial mastery against AI.

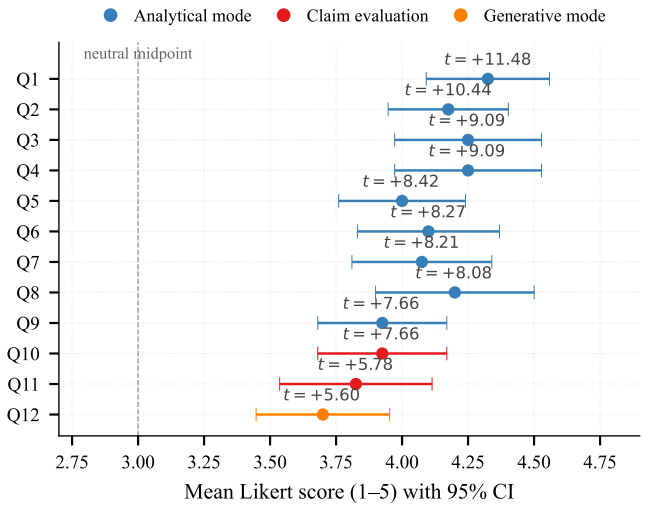

Figure 1: All 12 Likert items, *sorted top-to-bottom by descending t-statistic* against the neutral midpoint (3.0); all positive at $p < .001$. Analytical-mode items (blue) anchor the top; claim evaluation (red) and the generative-mode item Q12 (orange) anchor the bottom. Q-item text in Appendix A.

Implications. Three implications for the field. *Assessment shift:* protocol-survey production tasks (describe TCP, trace CSMA/CA) are AI-trivializable; assessments should move to framework-mediated claim evaluation tasks on NEW scenarios, including non-networking systems. The midterm instrument used here, including a load-balancer scenario analyzed structurally, is one model. *Reusable evaluation protocol:* the within-subject prior-course comparison (Q1, Q7) is reproducible by any instructor with a protocol-survey prerequisite, enabling cross-institution accumulation of evidence. *Portable research question:* the analytical-generative

asymmetry (§4.4) generalizes beyond networking; any first-principles framework can be tested on whether its categories support both decomposing instances and predicting from constraints. We expect the asymmetry to recur in disciplines that have taken similar pedagogical turns. The verification-heavy profile may be intrinsic to a decomposition-first framework; a complementary generative scaffold such as Clark's goal-ordering argument [4], deriving an architecture from a priority ordering over design goals, is a natural next step. Two questions remain open: whether a single introductory course can be refashioned beyond the layered structure to build these capabilities, and what assessment looks like when claim evaluation, not production, is the target.

REFERENCES

[1] Paul Baran. 1964. On Distributed Communications Networks. *IEEE Transactions on Communications Systems* 12, 1 (1964), 1–9.

[2] Giuseppe Bianchi. 2000. Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications* 18, 3 (2000), 535–547.

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)* (2020).

[4] David D. Clark. 1988. The Design Philosophy of the DARPA Internet Protocols. In *Proc. ACM SIGCOMM*. 106–114.

[5] Jim Gettys and Kathleen Nichols. 2012. Bufferbloat: Dark Buffers in the Internet. *Commun. ACM* 55, 1 (2012), 57–65.

[6] Arpit Gupta. 2026. Computing Is a Generative Discipline. Blog post, https://sites.cs.ucsb.edu/~arpitgupta/blog/.

[7] Arpit Gupta. 2026. A First-Principles Approach to Networked Systems. (2026). Working manuscript, https://sites.cs.ucsb.edu/~arpitgupta/first-principles-networking/.

[8] Arpit Gupta. 2026. Systems for Agents, Agents for Systems. Blog post, https://sites.cs.ucsb.edu/~arpitgupta/blog/.

[9] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A Buffer-Based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service. In *Proc. ACM SIGCOMM*.

[10] IEEE. 1997. *IEEE 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.* Technical Report. IEEE Standards Association.

[11] IEEE. 2021. *IEEE 802.11ax: Enhancements for High Efficiency WLAN.* Technical Report. IEEE Standards Association.

[12] Van Jacobson. 1988. Congestion Avoidance and Control. *ACM SIGCOMM Computer Communication Review* 18, 4 (1988), 314–329.

[13] James F. Kurose and Keith W. Ross. 2022. *Computer Networking: A Top-Down Approach* (8th ed.). Pearson.

[14] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, et al. 2017. The QUIC Transport Protocol: Design and Internet-Scale Deployment. In *Proceedings of the ACM SIGCOMM Conference*. 183–196.

[15] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. 2017. Neural Adaptive Video Streaming with Pensieve. In *Proc. ACM SIGCOMM*.

[16] Yakov Rekhter, Tony Li, and Susan Hares. 2006. *A border gateway protocol 4 (BGP-4).* Technical Report.

A LIKERT SURVEY ITEMS

The 12 five-point Likert items used in Figure 1 and discussed throughout §4. The scale runs from 1 (strongly disagree) to 5 (strongly agree). Items are renumbered in the order they appear in the figure (top to bottom by descending t-statistic against the neutral midpoint), so the same Q-number identifies the same item across the figure, the prose, and this appendix.

Q1 Compared to my prior networking course, this course gave me deeper understanding of WHY systems are designed the way they are.

Q2 The Environment-Measurement-Belief decomposition helped me understand WHY systems fail, not just that they fail.

Q3 Tracing how each generation of a technology solved the previous generation's failure (e.g., ALOHA → CSMA → CSMA/CA → OFDMA) helped me understand WHY each system was designed the way it was.

Q4 The lecture slides were structured so that I was asked a question BEFORE being given the answer. This helped me engage actively rather than passively listening.

Q5 I noticed that different systems (e.g., WiFi and cellular, TCP and ABR, scheduling and medium access) follow the same structural patterns when they face similar constraints.

Q6 The midterm tested my ability to apply the framework to NEW scenarios (TCP in a data center, concert venue WiFi, load balancer) rather than asking me to recall facts from lectures.

Q7 Compared to my prior networking course, I am better able to analyze a system I have NEVER seen before.

Q8 The lecture notes were detailed enough that I could understand the material by reading them, even if I missed the in-class lecture.

Q9 The four invariants (State, Time, Coordination, Interface) gave me a systematic way to analyze networked systems that I did not have before this course.

Q10 The first-principles framework makes me better at evaluating claims about systems, whether those claims come from AI tools, textbooks, or other people.

Q11 When I use AI tools (ChatGPT, Claude) for networking questions, the framework vocabulary (invariants, E-M-B, binding constraints) helps me formulate better questions and evaluate the answers more critically.

Q12 I can now predict how a system's architecture should look if I know its binding constraint (e.g., shared medium → distributed coordination, licensed spectrum → centralized scheduling).