FROM ENTROPY TO INTELLIGENCE.
RETHINKING SPECTRAL EFFICIENCY WITH AI
IN RAN.

Dr. Kim Kyllesbech Larsen

BOWW 2025 – Berlin.

10. September, 2025.

TechNEconomY


BPSK
BPSK
BPSK
BPSK
(very few passengers per car)
16-QAM
16-QAM
16-QAM
16-QAM
(higher number of passengers per car)
256-QAM
256-QAM
256-QAM
256-QAM
(high number of passengers per car)
Poor Road (SNR) – Limited Capacity
Good Road (SNR) – Higher Capacity
Great Road (SNR) – High Capacity

(Shannon's*) IDEAL WORLD

$$SE_{ideal} = \log_2(1 + SNR)$$

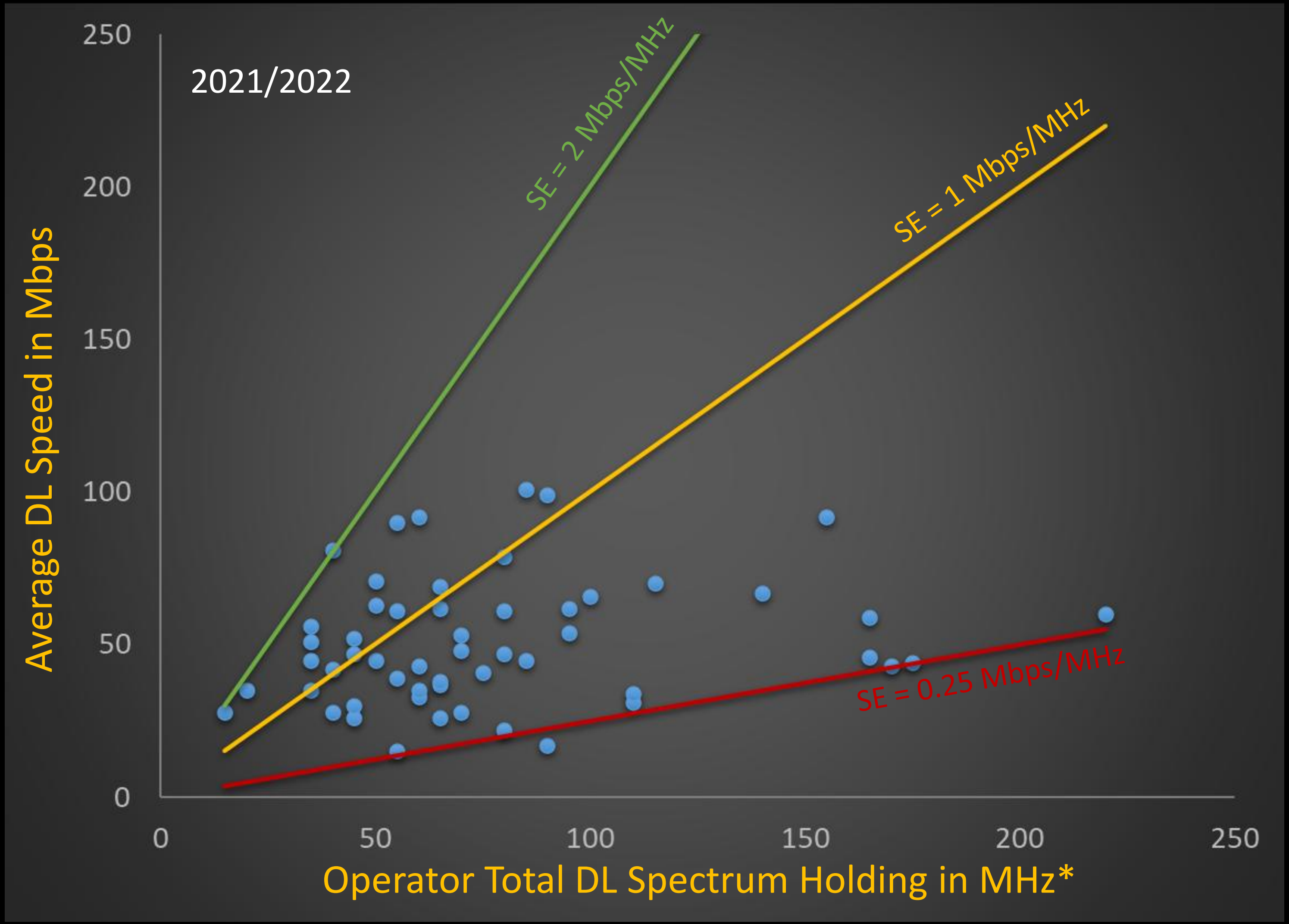
(Radio Engineer's) REAL WORLD

$$SE_{real} = (1 - O_{eff}) \cdot \log_2(1 + SINR)$$

(*) C.E. Shannon, "The Mathematical Theory of Communications" (1949), University of Illinois Press.

The spectral efficiency gap is evident. Operators are leaving Mbps and MHz in the drawer.

WEU Mobile Operators' Average DL Speed vs Their Spectrum Holding



Performance Ceiling.

- 4G LTE
  - Commercial LTE MiMo deployments rarely exceed 3 – 4 bps/Hz in real-world average performance.
  - LTE 8T8R passive MiMo has a max. **Shannon limit of** ~50+ bps/MHz (w. 8 layers).
- 5G NR
  - Commercial 5G NR massive MiMo deployments rarely exceed 5 – 7 bps/Hz in real-world average performance, with 3 - 5 bps/Hz in loaded networks.
  - 5G NR active mMiMo has a max. **Shannon limit** of 100s bps/MHz (12 DL layers ~256-QAM).

(*) Not including C-Band and mmWave possessions.

In the Real World, AI can boost the spectral efficiency by reducing inefficiencies in overhead (O) effects and increasing the SINR.

$$SINR = \alpha_{AI} \cdot \frac{P_s}{k_B T \cdot B + P_i^{\downarrow} + P_{res}^{\downarrow} + P_{imp}^{\downarrow}}$$

AI increases the Perceived spectral efficiency by reducing the relative entropy. ↑

↑ **AI improves the Quality of the Signal and increases the SINR.**

$$(P)SE_{eff} = \left[\frac{1}{\eta}\right] (1 - O_{eff}) \cdot \log_2(1 + SINR)$$

↓ **AI reduces the Overhead by optimizing the RAN Stacks' imperfections**

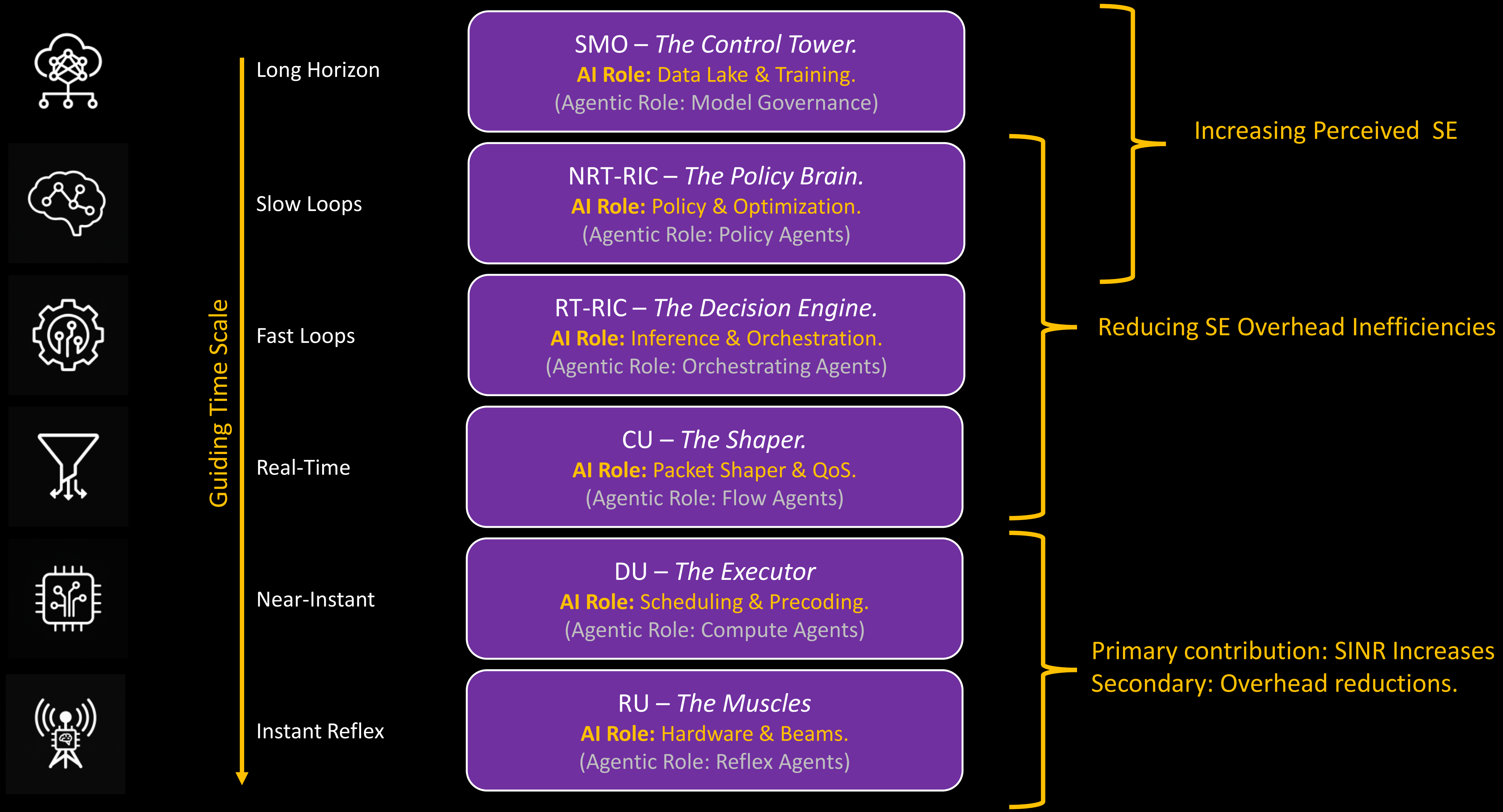
$$O_{eff}^{\downarrow} = 1 - (1 - O_{PHY}^{\downarrow}) \cdot (1 - O_{MAC}^{\downarrow}) \cdot (1 - O_{NET}^{\downarrow}) \cdot (1 - O_{TRA}^{\downarrow}) \cdot (1 - O_{APP}^{\downarrow})$$

PHYSICAL (L1) **Inefficiencies:** Channel estimation errors, Pilot density, Beam misalignment, HW impairments, …	SINR ↑ + O_{PHY} ↓
LINK LAYER (L2) **Inefficiencies:** HARQ retransmissions **(MAC)**, scheduling overhead, random access collisions, padding,…	O_{MAC} ↓ (SINR ↑)
NETWORK (L3) **Inefficiencies:** Routing/path inefficiency, IP fragmentation, Header overhead, Control chatter, …	O_{NET} ↓
TRANSPORT (L4) **Inefficiencies:** Flow control stalls, Retransmissions, ACK overhead, Congestion control oscillations, …	O_{TRA} ↓
APPLICATION (L5-L7) **Inefficiencies:** Semantic overfetch (e.g. unused video segments), Redundant data, No semantic compression, Inefficient encoding, Lack of compression, Verbose formats, Excessive session setup/teardown, Handshakes, State resynchronization, …	η ↓ (i.e., PSE ↑)

P_s: Received signal power at the receiver's antenna, **P_i**: Received interference power (not wanted) from other transmitters, **P_{res}**: Residual noise from imperfect channel estimation, quantization noise, and residual inter-symbol interference, **P_{imp}**: Hardware impairment noise from imperfections in the transceiver hardware. kB T: is the thermal noise power, and α_{AI} is the AI gain factor.

Brains, Decisions, Muscles: AI's Chain of Command in O-RAN.

Illustration.



SMO: Service Management & Orchestration, **NRT-RIC:** Non-Real-Time Radio Intelligence Controller, **RT-RIC:** Real-Time RIC, **CU:** Centralized Unit, **DU:** Distributed Unit, and **RU:** Radio Unit.

Automation (Today)

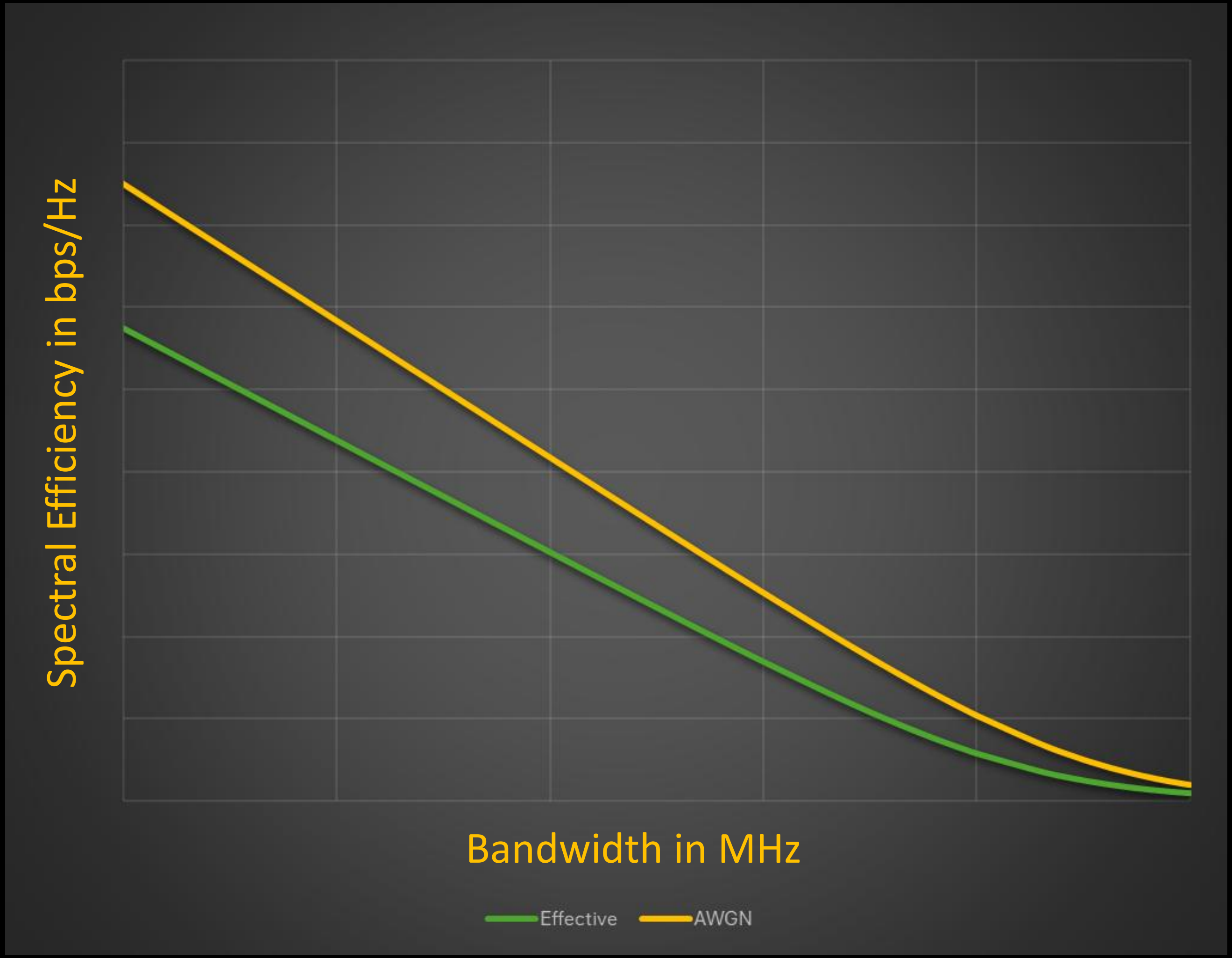
Agency (Near-Future)

Automaton vs **Agent**

Specialized functions, **fixed logic**, **narrow scope**, **static**, and **deterministic**.

Autonomous software that **perceives**, **decides**, **acts**, **learns**, and **adapts**. Not fully deterministic.

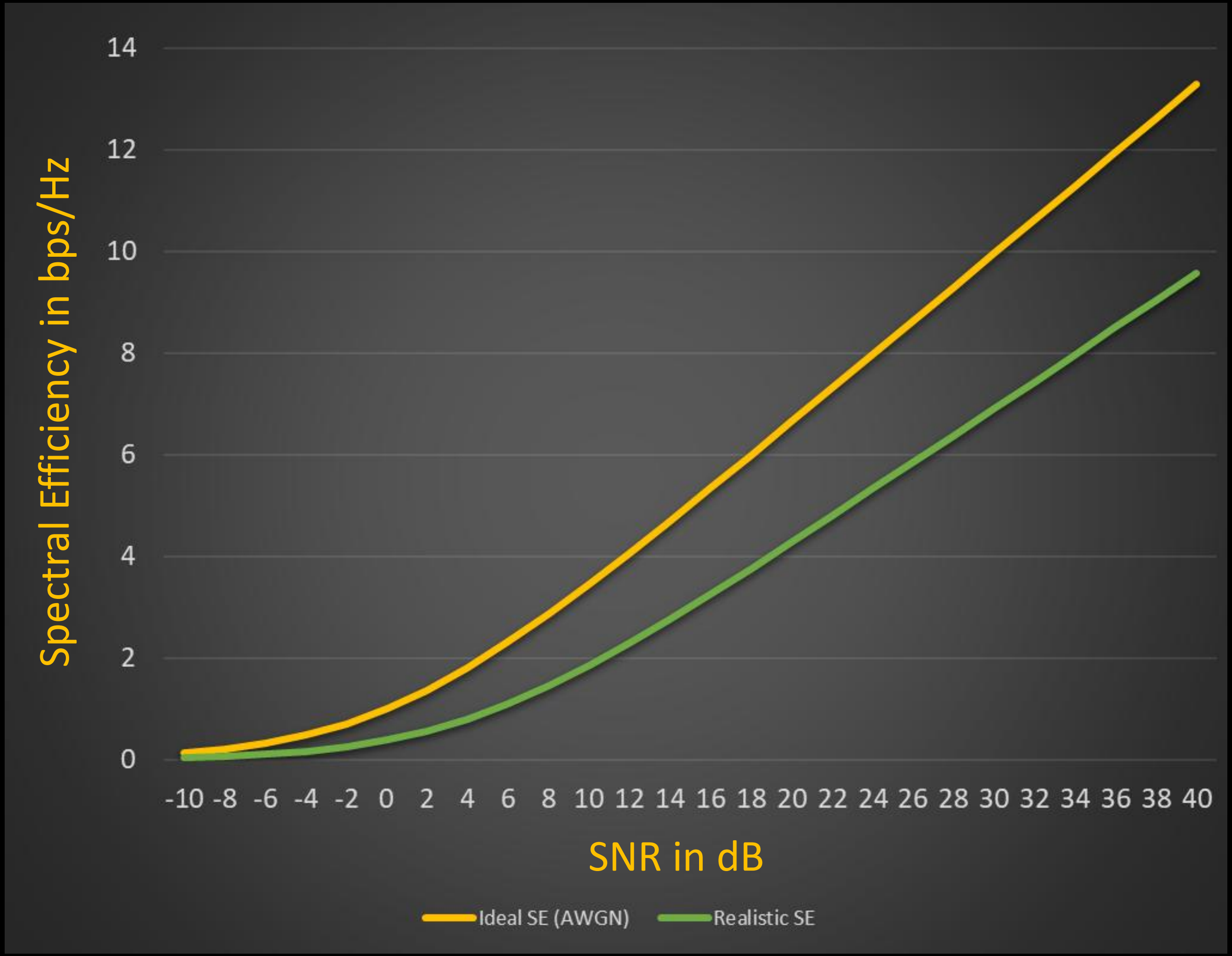
The Bandwidth Paradox: More Isn't Always Better, as spectral efficiency worsens as the bandwidth increases.



- Spectral Efficiency worsens as the Bandwidth Increases.
- Can be compensated by increasing the power (i.e., keep the spectral power density constant).
  - Works well for smaller bandwidth increases but has limitations.
- Typically, compensated by higher frequencies, increased bandwidth, and increased scale of massive MiMo antenna systems.
  - High frequency does lead to reduced SE (although not included in Shannon's work).

Note: AWGN = Additive White Gaussian Noise.

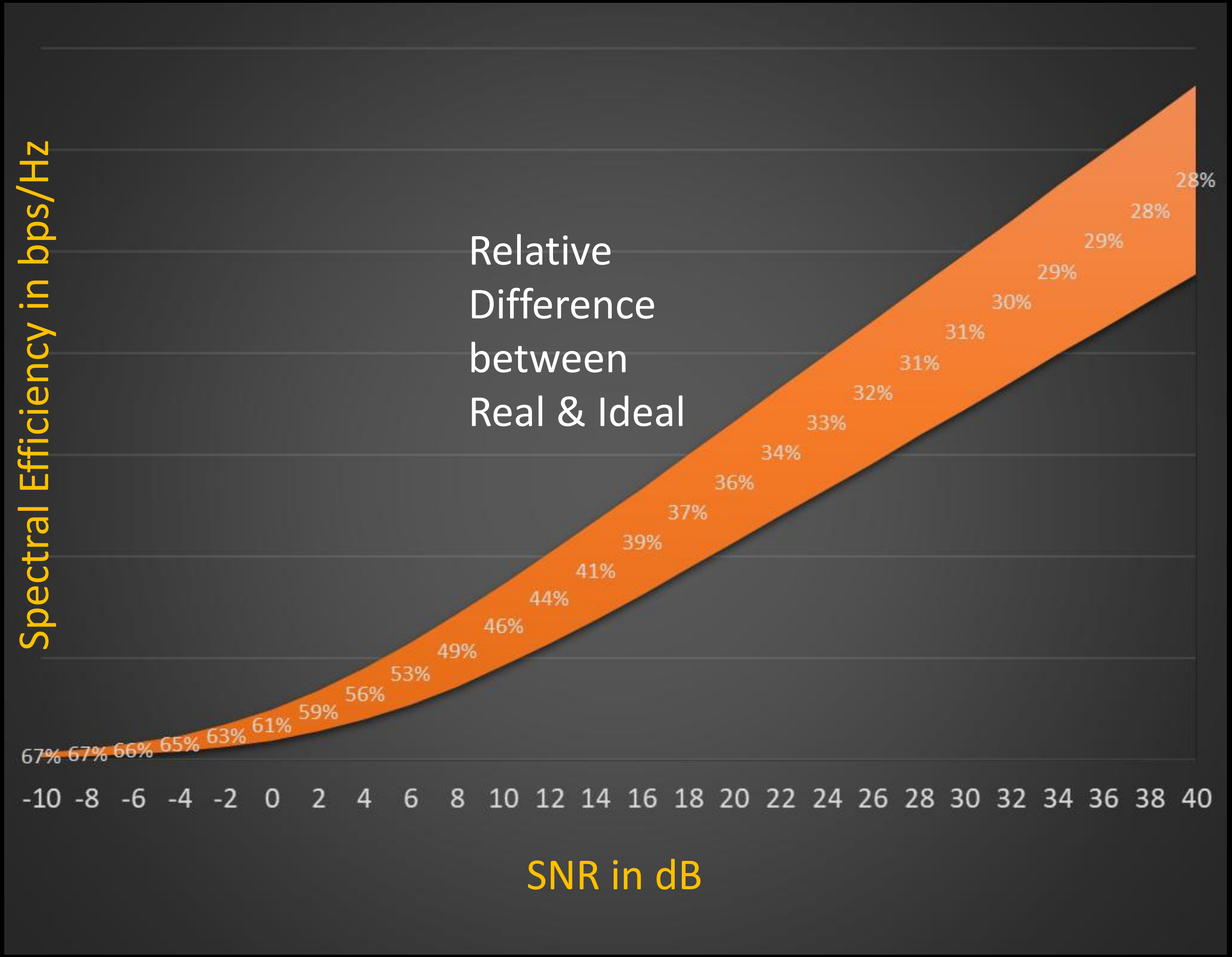
The difference between the ideal and the not-so-ideal is what we are hunting.



- The Shannon limit.
  - The absolute ceiling for spectral efficiency under ideal AWGN* conditions.
  - No real-world system can sustainably exceed it.
- Improvements to real-world
  - The green curve shows practical constraints like interference, fading, protocol overheads, hardware imperfections,…
  - AI and advanced techniques aim to narrow the gap to the yellow curve, not surpass it.

Note: AWGN = Additive White Gaussian Noise.

The relative penalty of non-ideal conditions is worst at poor SNRs (cell edge, deep indoor), while at high SNRs, the penalty is mostly an additive gap and not catastrophic.



- **Low SNR regime:** Noise-limited → small capacity easily wiped out by overhead.
- **High SNR regime:** Signal-rich → imperfections act like a fixed tax, relative losses are small.

Across DU, RIC, and RU, AI consistently squeezes more bits out of every Hz making up for channel and hardware imperfections.

O-RAN Layer	AI Role	SE Gain	Evidence Type	Notes
DU (MAC/PHY)	AI-native link adaptation.	+10% SE, +20% DL throughput.	Operator field trial (Ericsson + Bell).	Real-time (RT) inference in DU scheduler.
RT-RIC → DU	xApps for scheduling & optimization.	Up to +20% SE.	Research survey (O-RAN AI studies).	Policy-driven AI loops to DU.
DU (Scheduler)	DRL-based massive MIMO scheduling.	~+20% sum-rate.	Peer-reviewed (ArXiv/TMLCN).	Reinforcement learning (RL) scheduler.
DU (Testbed)	AI channel/MCS awareness + scheduling.	≥ +30% throughput.	Prototype testbed (academic/industry).	RT AI at DU; stronger channel prediction.
RT-RIC (Commercial)	Cohere USM (MU-MIMO & channel prediction).	Up to ~2× SE.	Vendor trial reports.	Runs as xApp; validation pending.
RU/DU (mmWave)	AI beam selection (sub-6GHz assisted).	−79.3% sweep overhead → higher SE.	Prototype (IEEE Access, OTA).	Improves SINR, reduces training overhead.

SMO: Service Management & Orchestration, **NRT-RIC:** Non-Real-Time Radio Intelligence Controller, **RT-RIC:** Real-Time RIC, **CU:** Centralized Unit, **DU:** Distributed Unit, and **RU:** Radio Unit.

AI addresses inefficiencies in modulation, scheduling, interference, and overhead using existing CQI, CSI, and HARQ data, deployable as rApps and xApps with immediate spectral efficiency gains.

- AI Enables Smarter Link Adaptation (DU).
  - AI predicts the optimal Modulation & Coding Scheme (MCS) better than rule-based algorithms.
  - Cuts HARQ retransmissions → unlocks higher-order modulation more often.
- AI Enables Interference-Aware Scheduling (DU / RT-RIC).
  - AI-powered schedulers allocate resource blocks while accounting for inter-cell interference.
  - Improves SINR → higher spectral efficiency without extra spectrum.
- AI Improved Beamforming & CSI Prediction (RU/DU).
  - AI predicts channel state information and optimizes Massive MIMO beam selection.
  - Reduces pilot overhead and improves signal quality → higher bits/Hz.
- AI Enables Overhead Reduction in Control Signaling.
  - AI detects and reduces redundant retransmissions, padding, and protocol overhead.
  - More of the bandwidth carries payload instead of signaling.

CQI: Channel Quality Indicator, CSI: Control State Information, HARQ: Hybrid Automatic Repeat Request, MCS: Modulation & Coding Schemes.

THANK YOU!

Contact:
Email: kim.k.larsen@techneconomy.com
Linkedin: www.linkedin.com/in/kimklarsen
Blogs: www.aistrategyblog.com & www.techneconomyblog.com
Twitter: @KimKLarsen

AI can boost the spectral efficiency by reducing inefficiencies in overhead effects and in the impact on SINR.

$$\frac{1}{\eta}$$

AI increases the <u>Perceived</u> spectral efficiency by reducing the relative entropy, using, for example:

- Semantic extraction.
- Content regeneration.
- App-aware video shaping.

$$SINR = \alpha_{AI} \cdot \frac{P_s}{k_B T \cdot B + P_i^{\downarrow} + P_{res}^{\downarrow} + P_{imp}^{\downarrow}}$$

AI improves the Quality of the Signal and increases the SINR, by:

- Interference suppression.
- Beamforming & Array Gain.
- Channel Estimation & CSI Prediction.
- Mobility & Handover Prediction.
- HW Impairment Compensation.
- Power Control & Resource Allocation.
- Sub-band & Carrier Selection.
- ….

$$\text{(P)SE}_{\text{eff}} = \left[\frac{1}{\eta}\right]^{\uparrow} (1 - \text{O}_{\text{eff}}^{\downarrow}) \cdot \log_2(1 + \text{SINR}^{\uparrow})$$

$$O_{eff}^{\downarrow} = 1 - (1 - O_{PHY}^{\downarrow}) \cdot (1 - O_{MAC}^{\downarrow}) \cdot (1 - O_{NET}^{\downarrow}) \cdot (1 - O_{TRA}^{\downarrow}) \cdot (1 - O_{APP}^{\downarrow})$$

AI reduces the Overhead by optimizing, for example:

- Pilots (DMRS, CSI-RS), Beam training, Synchronization, Guard bands (Physical Layer O_{PHY})
- HARQ retransmissions, Scheduling grants, Control channel signaling (Medium Access & Control O_{MAC}).
- TCP/IP/QUIC Headers, Congestion control inefficiencies, head-of-line blocking (Transport/Network Protocol Overhead O_{NET} O_{TRA}).
- ABR video overfetch, unused prefetched segments, redundant media frames, … (Application Overhead & Waste O_{APP})

Note: Opanga RAIN AI Platform solution, see https://www.opanga.com/the-rain-overview, directly impacts APP, TRA, and NET overheads and can substantially boost the spectral efficiency (SE) in capacity limit communications systems, as well as improve the SE under poor SINR/SNR conditions.

Brains, Decisions, Muscles: AI's Chain of Command in O-RAN.

Illustration.

The AI Process:

Service Management & Orchestration (SMO) – *The Control Tower.*
AI Role: Data Lake, Policy Management & Machine Learning (ML) Training & Semantic levers.

SMO: Ingest → Labeling → Training → Approval (Human) → Curation, and Versioning. Publishes to the NRT-RIC.

Non-Real Time RAN Intelligent Controller (NRT-RIC) (rApps: > 1 s) – *The Policy Brain.*
AI Role: acting on longer time scales of a second or higher.
Policy computation, optimization, revision (sec, hours, days,…) & ML Models.

NRT-RIC: Deployment-ready Models (from SMO), Model Selection & finetuning. Further dispatches to RT-RIC.

Real-Time RIC (xApps: 10 ms – 1 s) – *The Decision Engine.*
AI Role (Today): Short-horizon inference & policy arbitration on telemetry; Carrier/Sub-band, CoMP & cross-cell coordination; safety/guardrails.
AI Role (Future): AI Agent Orchestration, responsible for composing, dispatching, monitoring, and rolling back small ***goal-driven agents*** that execute inside the DU (and, selectively, RU/MIMO fabric) under hard real-time budgets.

RT-RIC: Ingest telemetry → inference/predictions → control parameters → push to DU & CU.

Reducing SE Overhead Inefficiencies

Centralized Unit (CU)
AI Role: acting on time scales from 10 ms to 100s of milliseconds. The CU is not tied to TTIs.

….

CU – *The Shaper.*
AI Role: Packetization/MTU tuning, header compression, QUIC/TCP optimization

CU: RT-RICs parameters → CU populate models that act (10+ms) on packet handing & QoS flows, as well as on control loops.

SINR Increases & Reduce SE Overhead Inefficiencies

Distributed Unit (DU)
AI Role: acting on per-TTI* time intervals, scheduling/precoding, host tiny inline models, Power Control learning,

….

DU – *The Executor.*
AI Role: MiMo Compute fabric, Grouping, MCS & HARQ Prediction, CSI prediction, Carrier & sub-band Selection,

….

DU
AI Role: Pilot contamination mitigation, CoMP optimization, traffic-aware scheduling, …

DU: Takes the RT-RICs parameters → set constraints, targets, and weights acting on Upper-PHY, MAC, and RT-control loops.

Reducing SE Overhead Inefficiencies

Radio Unit (RU)
AI Role: acting on per-TTI* time intervals,

…

RU
AI Role: HW impairment compensation

RU
AI Role: Element outlier detection & de-weighting,

…

RU
AI Role: Beam weight & tracking,

RU

…

RU

RU: Takes the DUs parameters → set constraints, targets, and weights acts on a 1/10 - 1 ms level.

(*) TTI (Transmission Time Interval is 1 ms for 4G, and between 1 ms and 0.0625 ms depending on subcarrier spacing. Anything TTI-based is not for the RIC but happens on a DU and RU level.