Canova Tech

The Art of Silicon Sculpting

PIERGIORGIO BERUTO
ANTONIO ORZELLI

IEEE802.3cg TF

Proposal for short-reach multi-drop 10M SPE (former PLCA)

September 14th, 2017

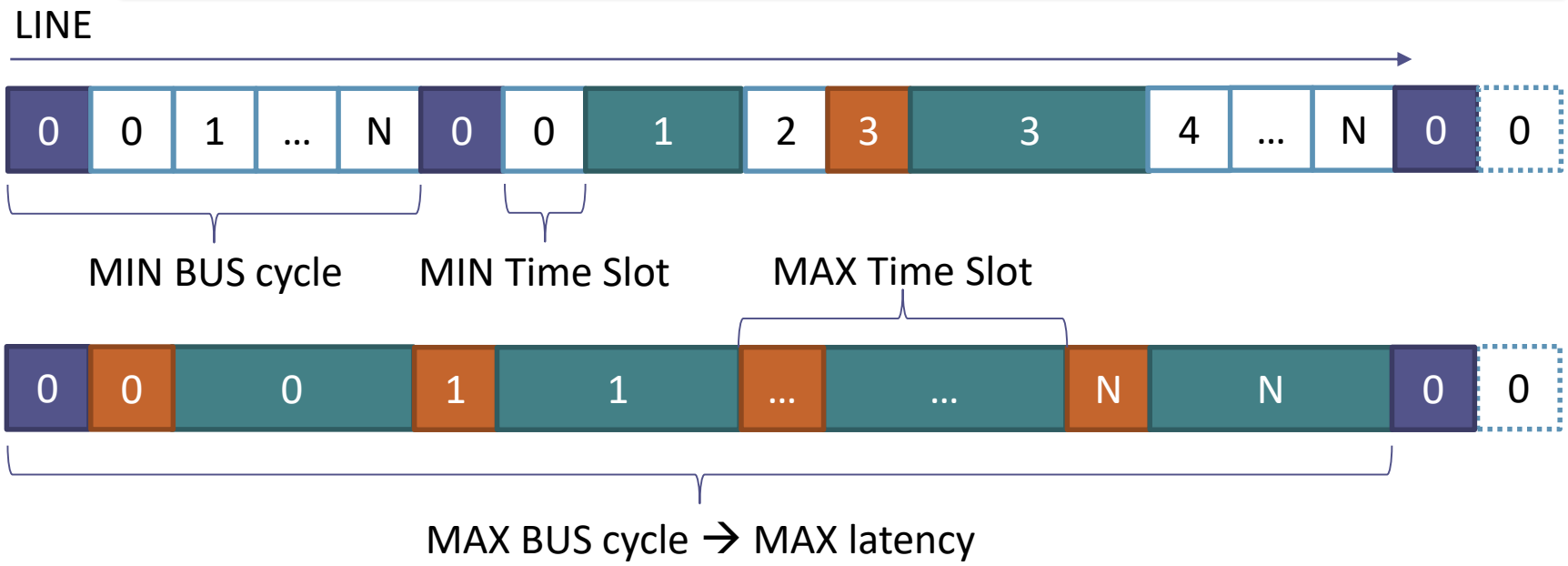
Background

- Proposal was called “PHY Level Collision Avoidance” – because it tries to avoid physical collisions on the line
  - This has caused some confusion
- Proposal has evolved to fit within normal CSMA/CD
  - No more explicit “COMMIT” or “YIELD” transmissions
  - “YIELD” is replaced with silence, as in normal Ethernet
  - “COMMIT” is replaced by transmitting data or idle, as in normal Ethernet
- Multi-access technique is Carrier Sense / Collision Detect
  - PHY can only assert carrier on the line during its prescribed time slot
  - Other PHYs detect collisions before actually putting data onto the line
  - Data is always sent within the second MAC TX attempt

Overview

- What it really is:
  - Proposal for half-duplex, multi-drop, short-reach 10M SPE PHY
  - Collision detection mechanism
    - Avoids physical collisions on the media (throughput)
    - Guarantees latency < NUM_PHY * MAX PKT LENGTH (fairness)
    - Transparent to upper layers
- Objectives
  - Interworking with standard CSMA/CD MAC
    - No modifications to MAC, everything done at PHY level
  - Beat plain CSMA/CD performance (throughput/latency/fairness) especially at high network loads
  - Keep complexity low
  - Support up to 8 nodes, possibly more
  - Shall work with different PCS coding

In a nutshell



- PHYs are statically assigned unique IDs [0..N]
- PHY with ID = 0 is the master
  - Sends BEACON to signal the start of a BUS cycle and let slaves synchronize their transmission processes
- A BUS cycle consists of N+1 **variable size** time slots plus the initial BEACON
  - PHYs are allowed to transmit only during the time slot which number matches their own ID
  - Time Slots end if nothing is transmitted within "MIN Time Slot" period or at the end of any transmission
  - PHYs are allowed to pad their own time slots with IDLE to compensate for any MAC latency (e.g. IFG)
- In numbers
  - BEACON time == MIN Time Slot == ~20 bT (bit-times)→ ~1.6µs assuming 4b/5b + DME encoding
  - MAX latency == BEACON time + PHYs LATENCY +MAX BUS Cycle (all PHYs transmit one packet of MAX size, e.g. 1542 Bytes including IFG) →~12.500 bT * (N+1) → ~1ms per PHY assuming 4b/5b + DME
- Round-robin scheduling guarantees fairness

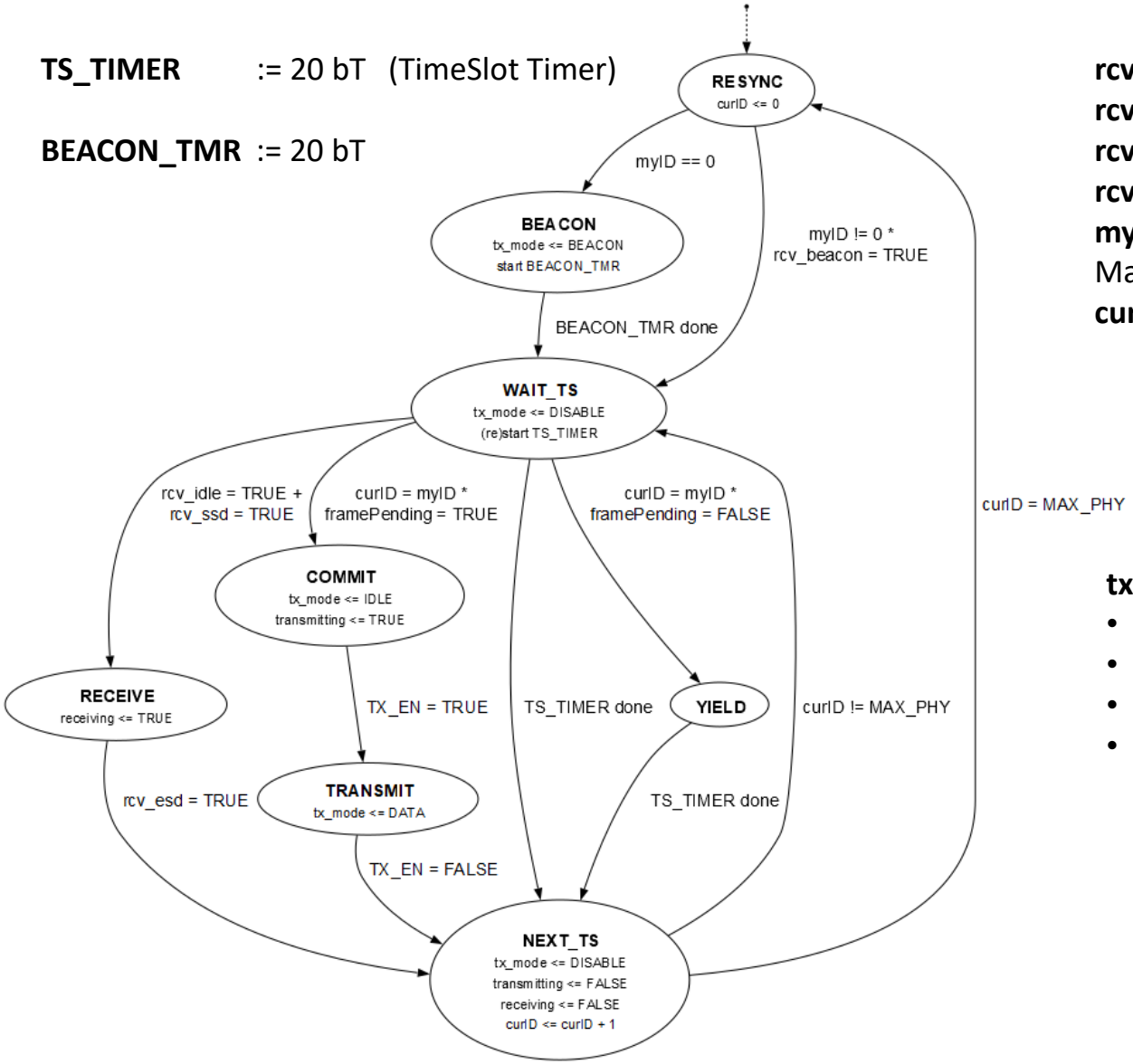
Definitions

- Collision Detect mechanism can be defined with two parallel processes
  - TX process
    - defines how PHYs synchronize to transmit data
  - COL process
    - defines how collision detection is reported to the MAC
- Shared variables:
  - framePending
  - transmitting
  - receiving

TX-PROCESS

TS_TIMER := 20 bT (TimeSlot Timer)

BEACON_TMR := 20 bT

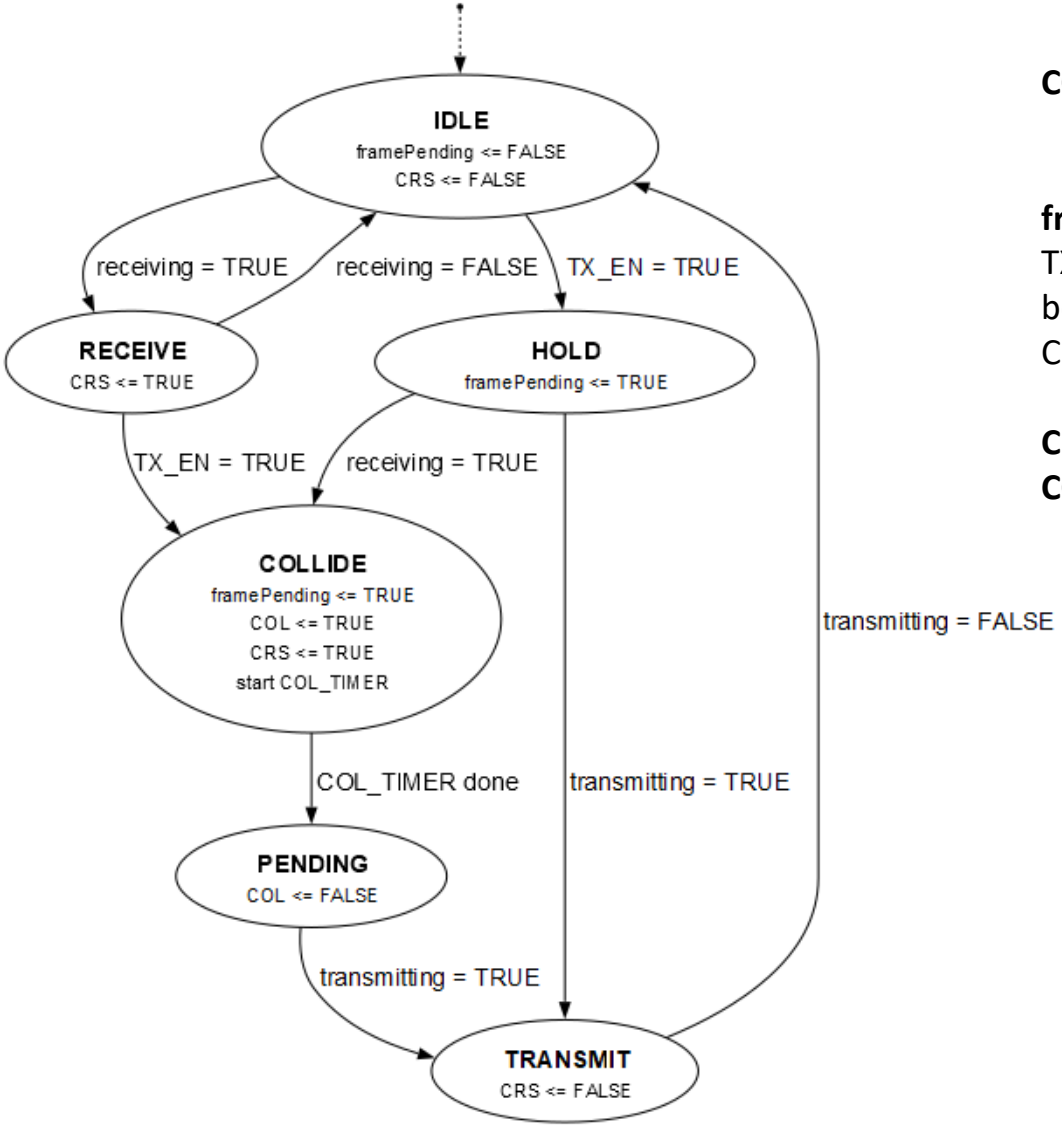


rcv_beacon := BEACON signal received
rcv_idle := IDLE signal received
rcv_ssd := SSD received
rcv_esd := ESD or ESD/ERR received
myID := PHY ID configured by Layer Management
curID := current Time Slot ID

tx_mode :=

- DISABLE: transmitter disabled
- IDLE: send IDLE pattern
- BEACON: send BEACON pattern
- DATA: send encoded transmit data

COL-PROCESS

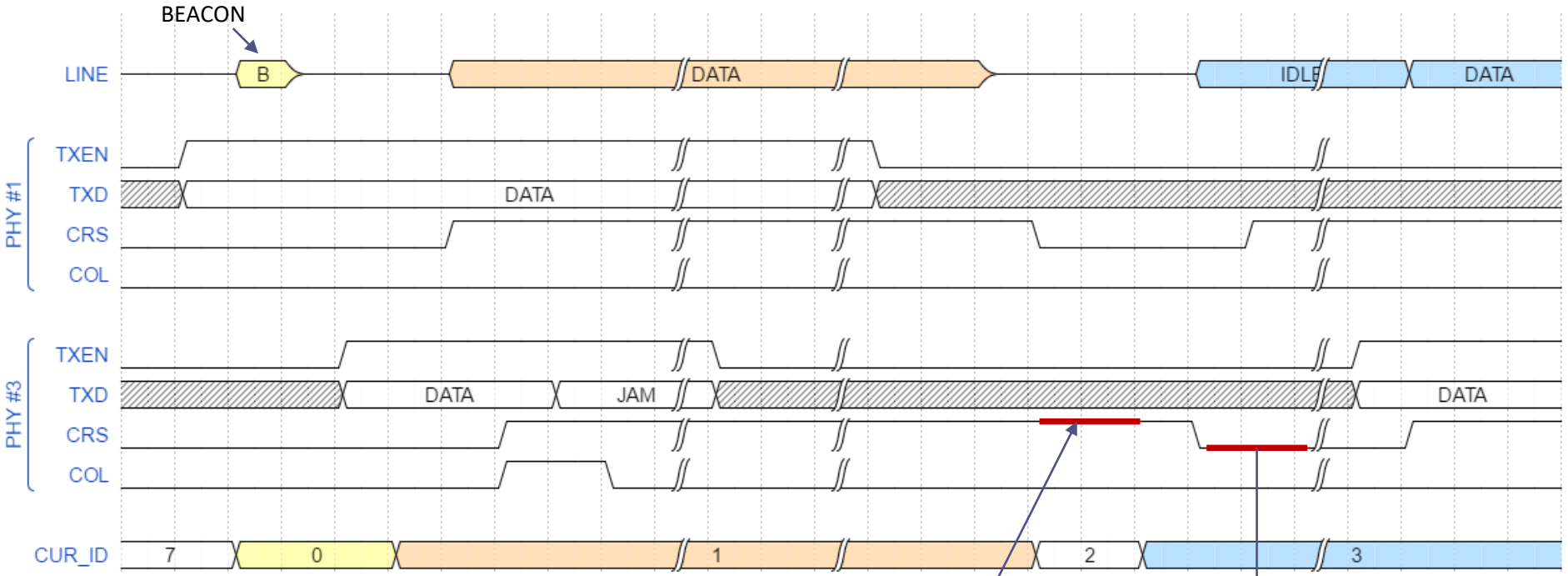


COL_TIMER := 10 bT

framePending := latched version of MII TX_EN. It is set as soon as TX_EN becomes TRUE and it's reset as soon as CRS transitions from TRUE to FALSE

CRS := Carrier Sense
COL := Collision

Example Waveforms

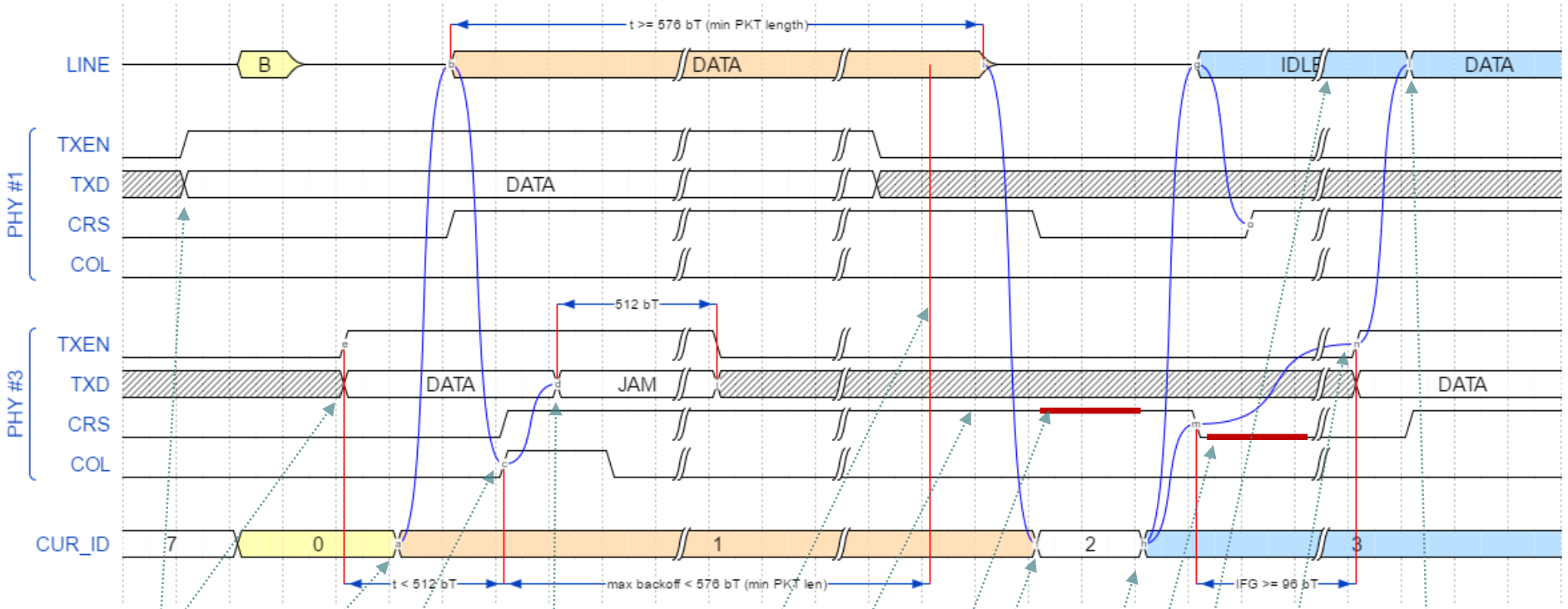


- BUS with 8 nodes
- Node #1 and #3 want to transmit data, others are silent
  - PHY #1 just defers TX until its own time slot is available
  - PHY #3 signals a collision because PHY #1 is transmitting, however:
    - No physical collisions on the line
    - Actual TX occurs immediately after PHY #1 transmission with no additional delay (MAX backoff + latency < MIN packet size)

CRS forced HIGH to prevent the MAC from transmitting until CUR_ID = 3

CRS forced LOW to have the MAC deliver the packet

Example Waveforms



MAC #1, 3 start transmitting.
PHY #1, 3 framePending <= TRUE

PHY #1 time slot begins, data is put on the line since framePending = TRUE

PHY #3 signals a collision to its MAC since PHY #1 carrier is sensed

MAC #3 initiates backoff and sends JAM in response

MAC #3 backoff time always ends before PHY #1 transmission is over (attempt = 1)

MAC #3 does not perform a new attempt (yet) because CRS is asserted

Time slot #1 ends when PHY#3 falls silent again

PHY #3 keeps CRS asserted as framePending == TRUE

Time slot #2 is yielded

PHY #3 de-asserts CRS (allow MAC to perform new attempt)

Phy #3 puts IDLE on the line to extend the time slot until MAC is transmitting

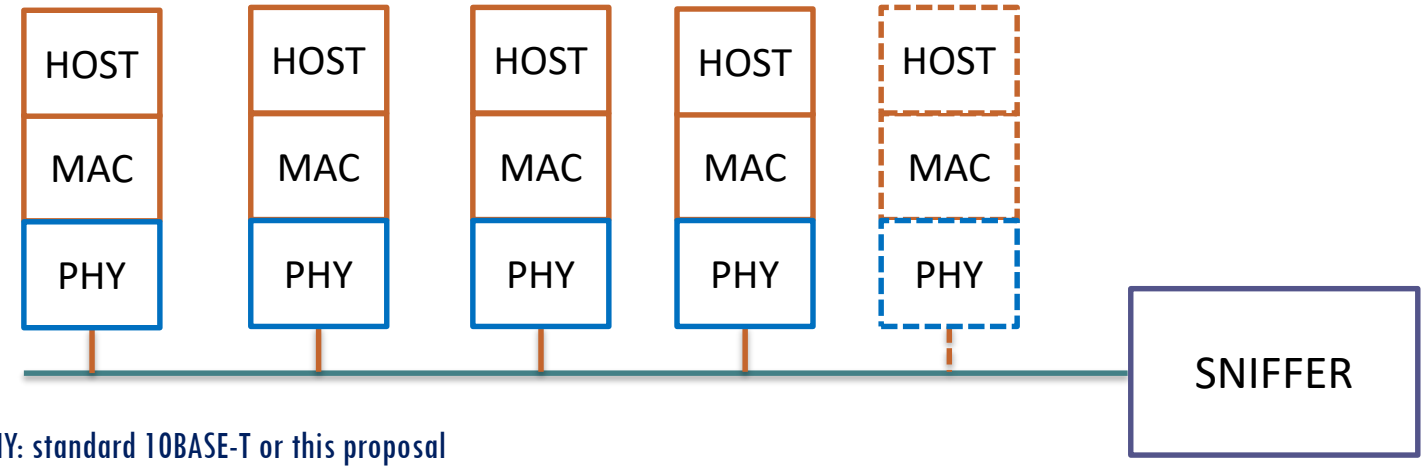
MAC #3 waits for IFG then attempts transmission again. DATA is eventually put on the line

Caveats & Constraints

- Each PHY keeps track of "curID" on its own, re-synchronizing the clock at each BEACON.
  - in real world, PHY TX/RX latency and line propagation delay cause clock skew across the PHYs
    - system won't break as long the carrier is sensed within the minimum time slot period, that is TS_TIMER.
      - we can live with clock skew by taking appropriate margin on TS_TIMER
      - TS_TIMER could be configurable via management interface anyway
  - In numbers
    - line propagation delay for a 25mt cable is ~150ns that is ~2 bit times assuming 4b/5b and DME
      - almost negligible
    - Setting TS_TIMER to 20 bT (~1.6μs) allows for a total TX to RX internal PHY latency of 8 bT
      - reasonable for most PCS and PMA implementation
- Any delay (due to specific implementation) between carrier sensed and collision asserted might cause backoff time to expire after the shortest allowed packet is actually sent on the line.
  - doesn't break the system but affects performance (longer IDLE / wasted time slots)
  - margin is 12 bT considering 20 bT of PHY latency and MAX backoff time
    - should be more than enough for most implementations

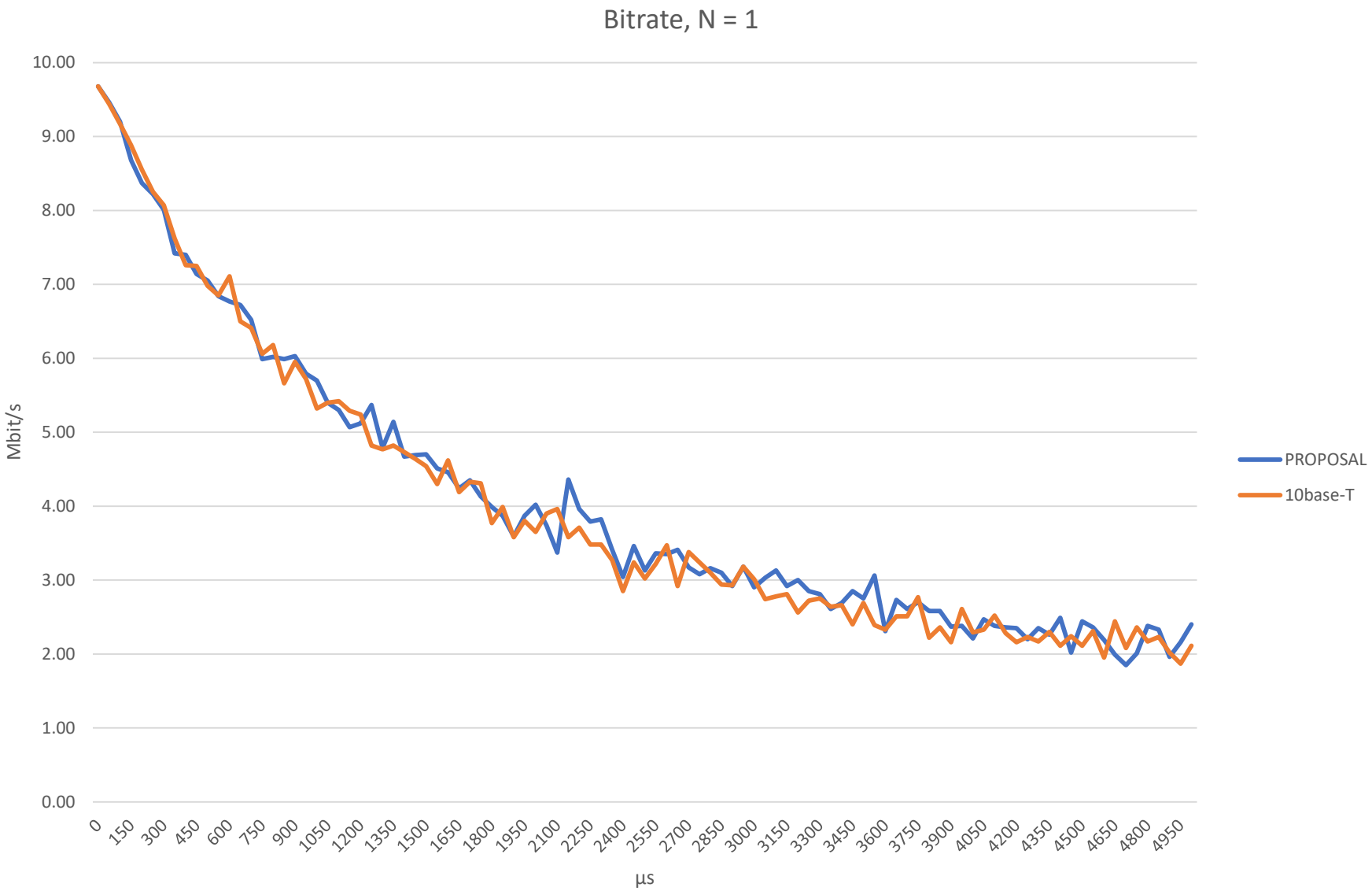
Simulations

- Full digital simulation (verilog)
  - 4b/5b encoding + DME

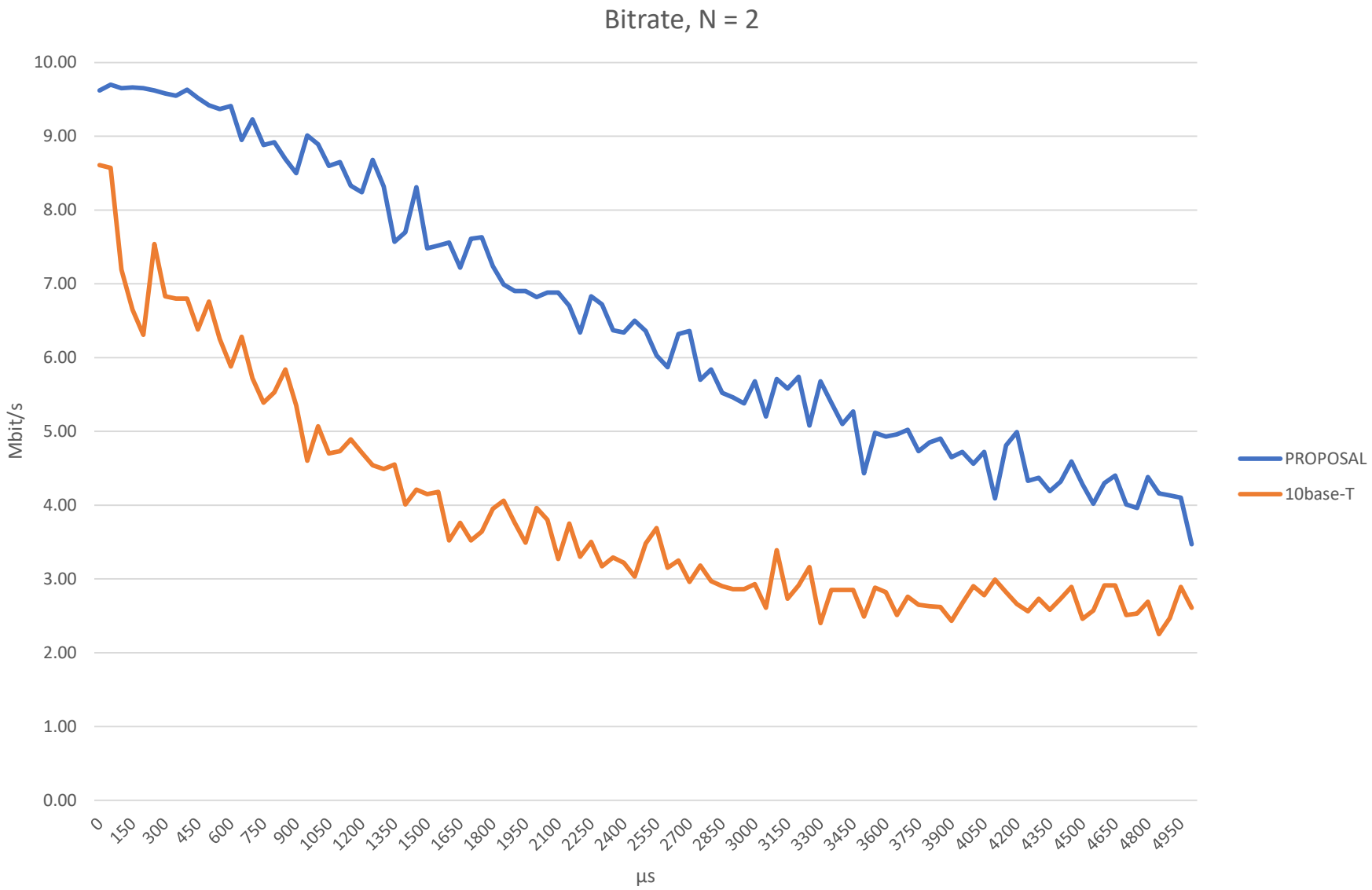


- PHY: standard 10BASE-T or this proposal
- MAC: standard CSMA/CD capable MAC (802.3 clause 4)
  - host interface: DPRAM (one frame) + busy indication + size + trigger
  - PHY interface: MII (txd, txclk, txen, txer, rxd, rxclk, rxdv, rxer, col, crs)
- HOST: simple transmitter
  - wait for MAC BUSY = 0
  - wait random time between 0 and MTP (sim parameter, 0 = MAX speed)
  - write random data in DPRAM of size PKTSZ (sim. parameter 60 < PKTSZ < 1500) or random size
- SNIFFER: measuring throughput, latency
  - throughput: number of received bytes (excluding FCS, PREAMBLE) / total simulation time
  - latency: time between MAC BUSY = 1 and MAC BUSY = 0 for each node

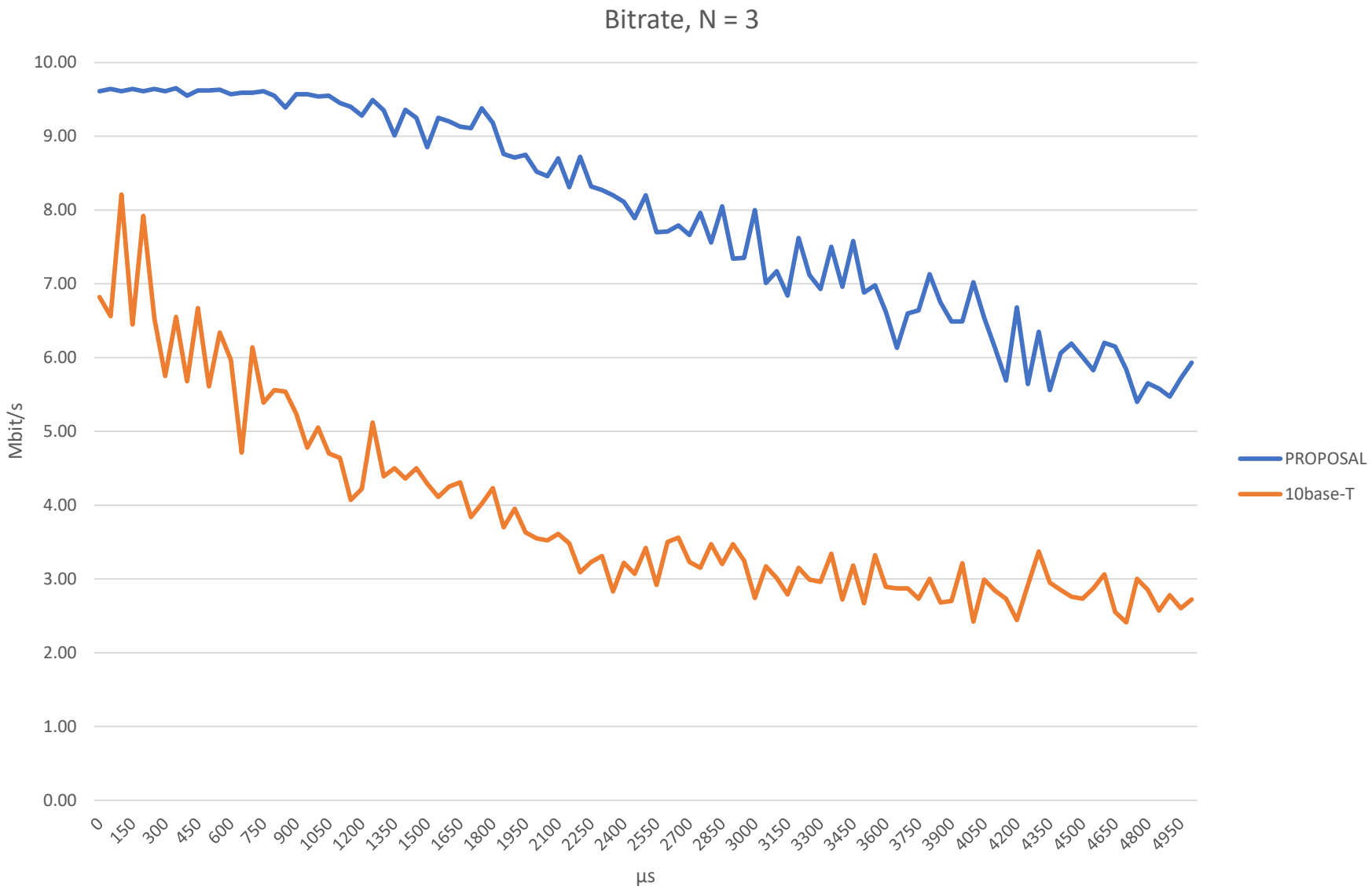
Simulations



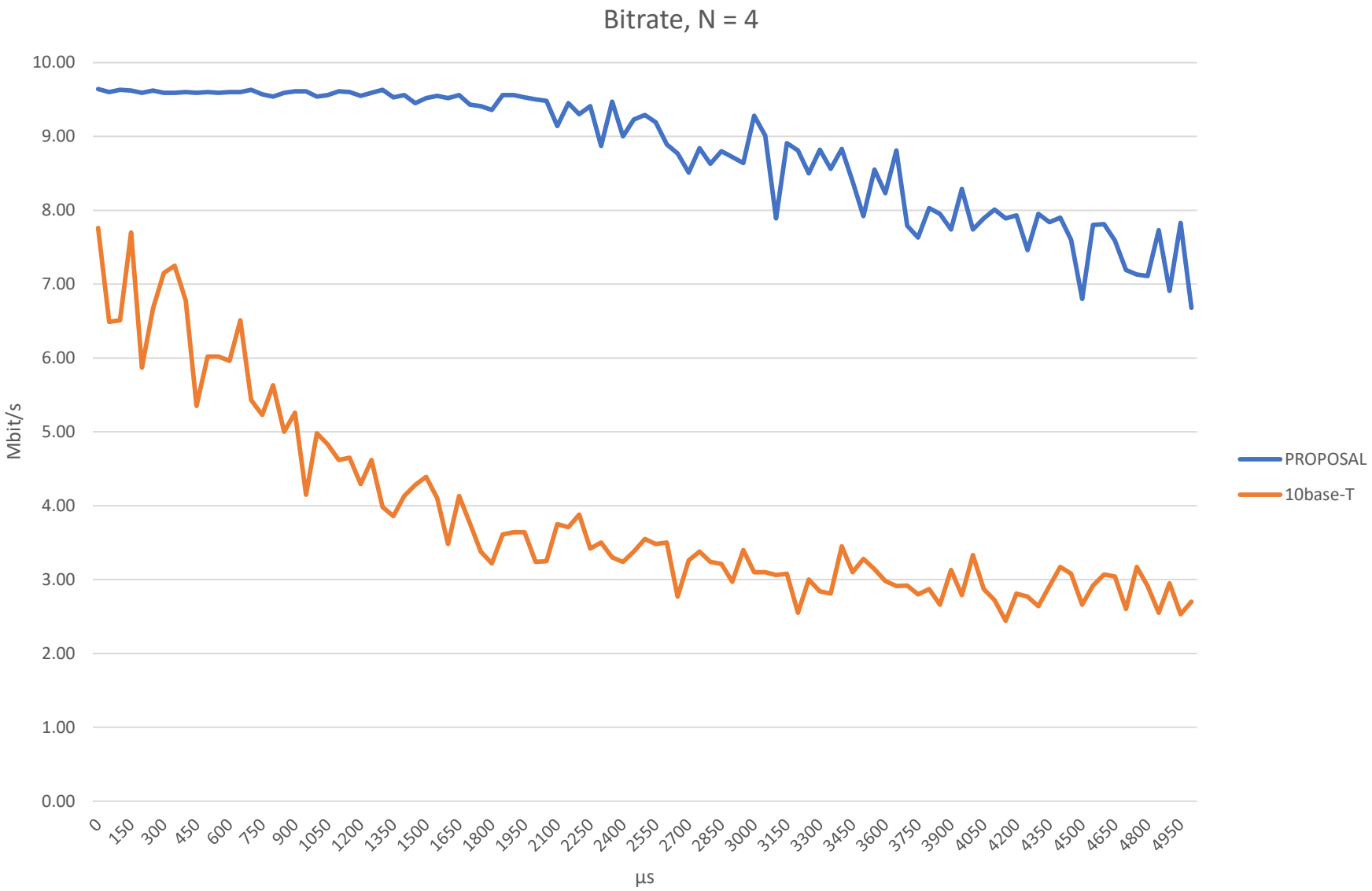
Simulations



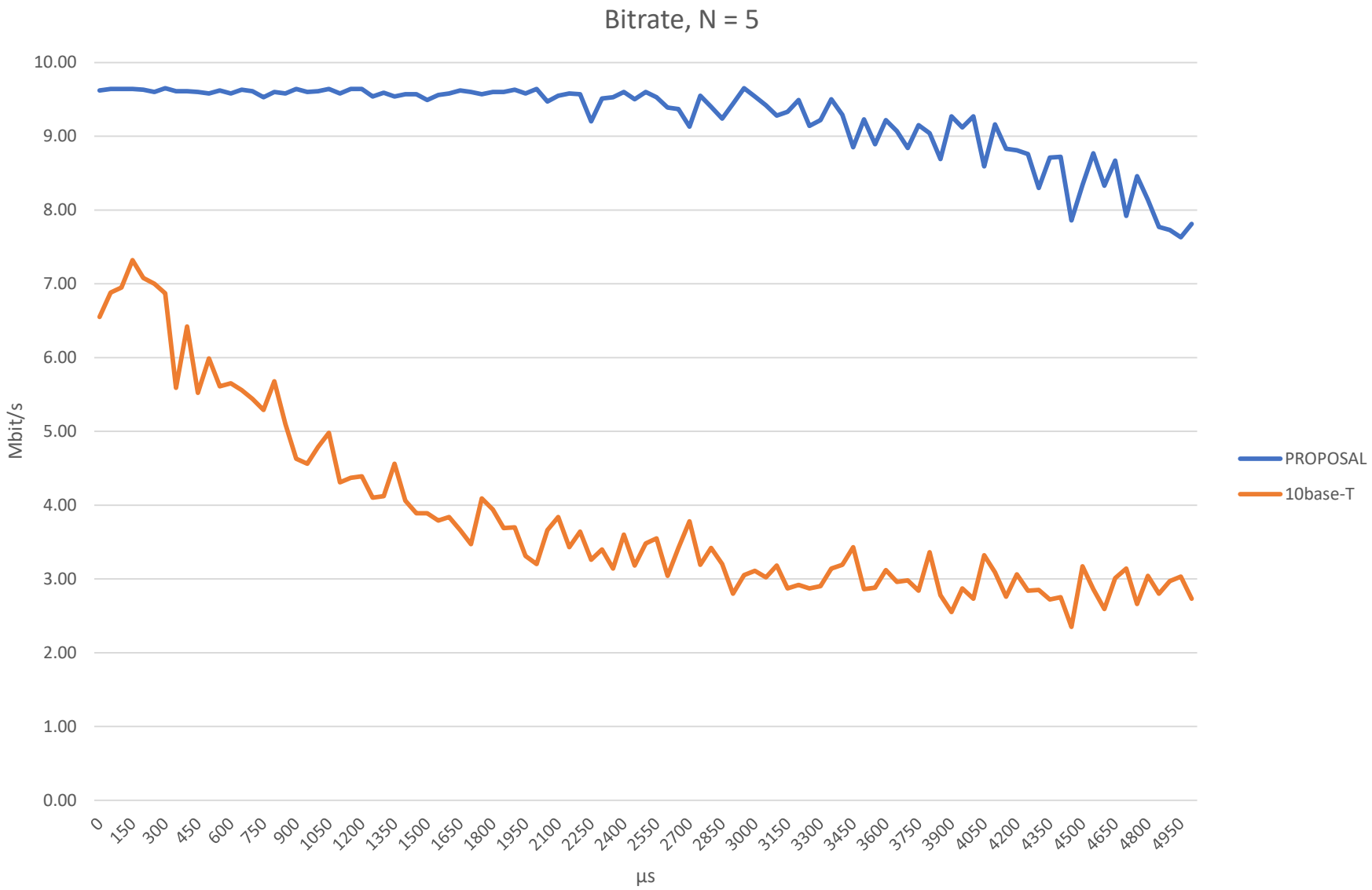
Simulations



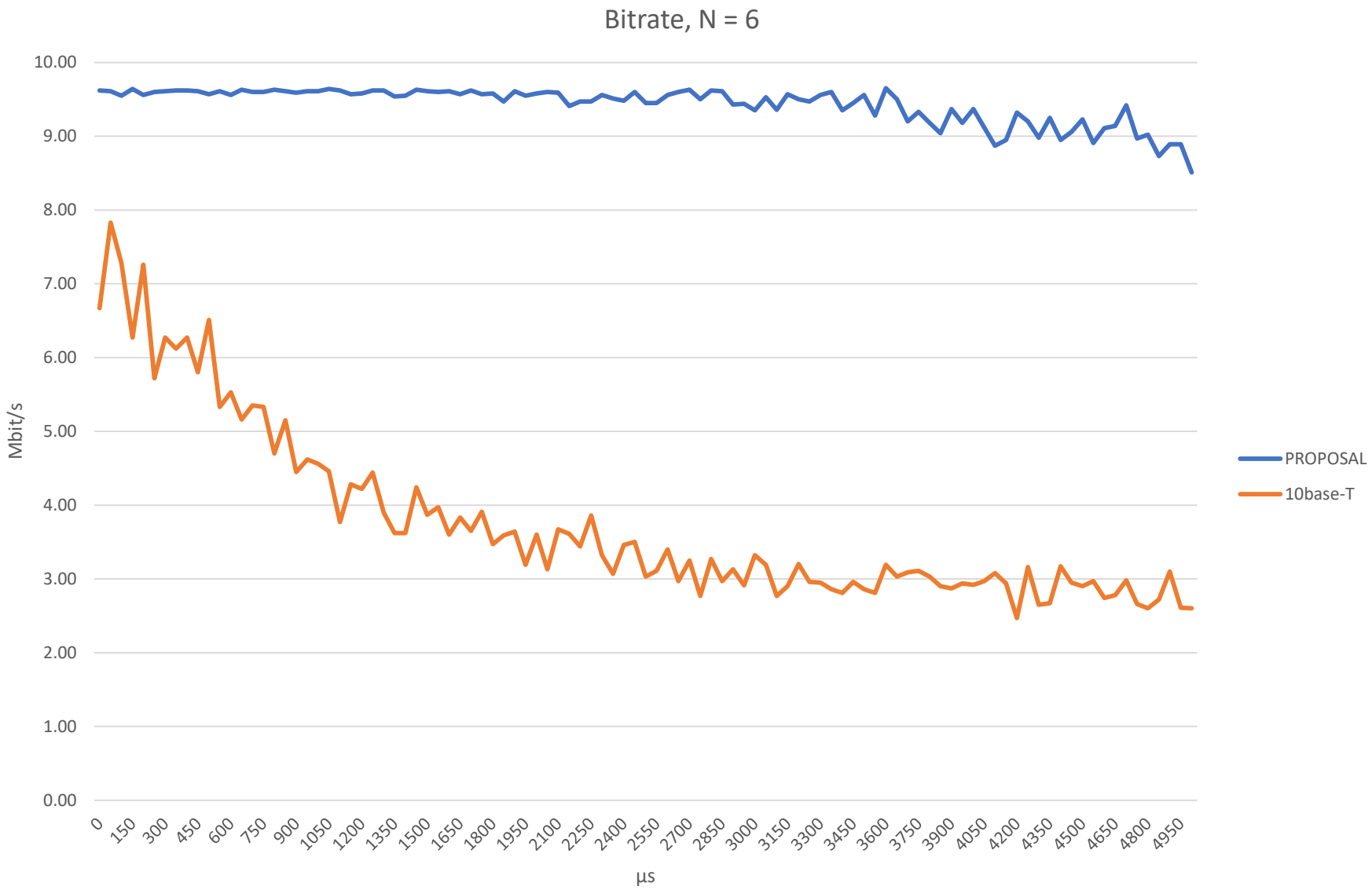
Simulations



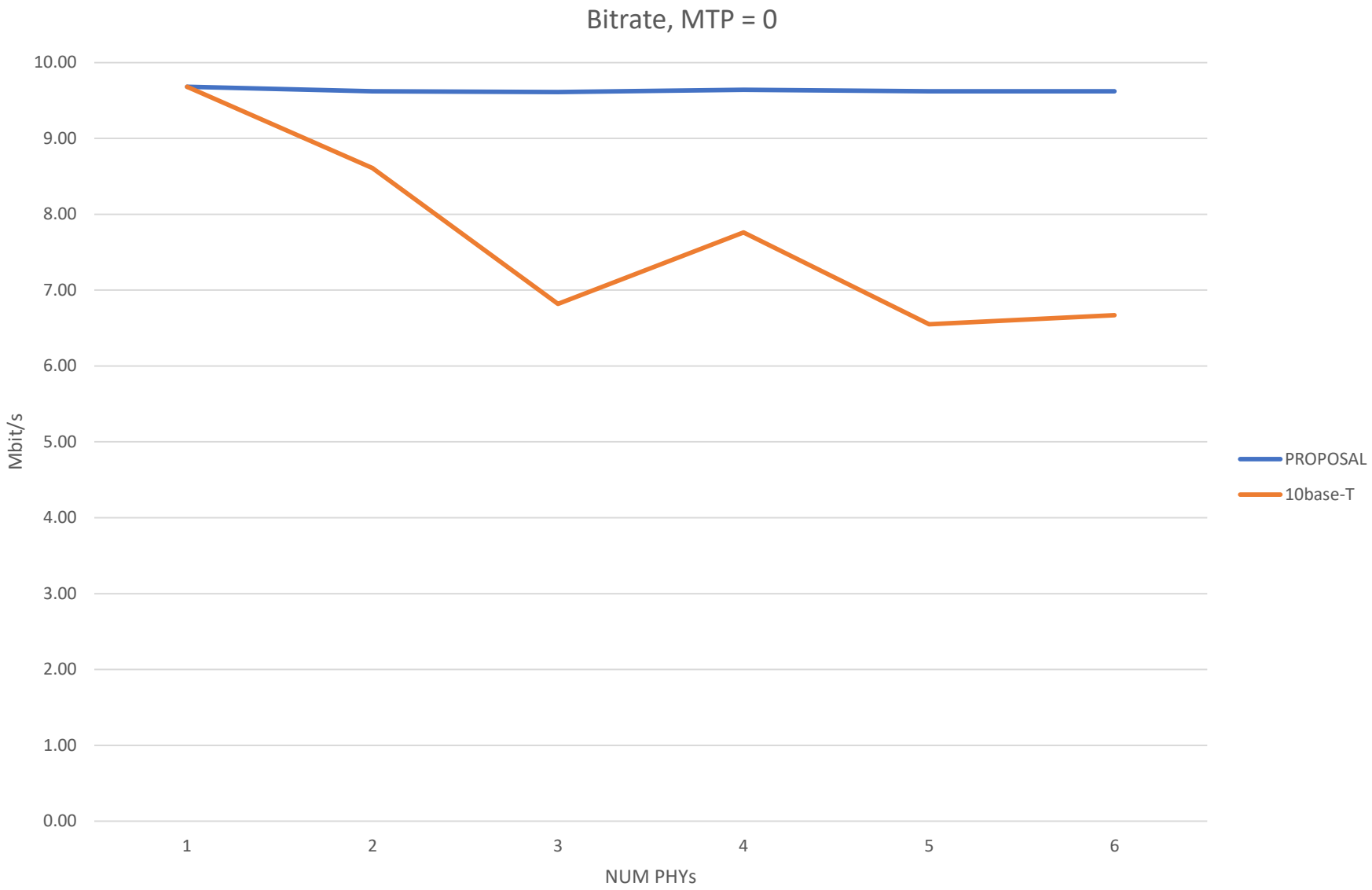
Simulations



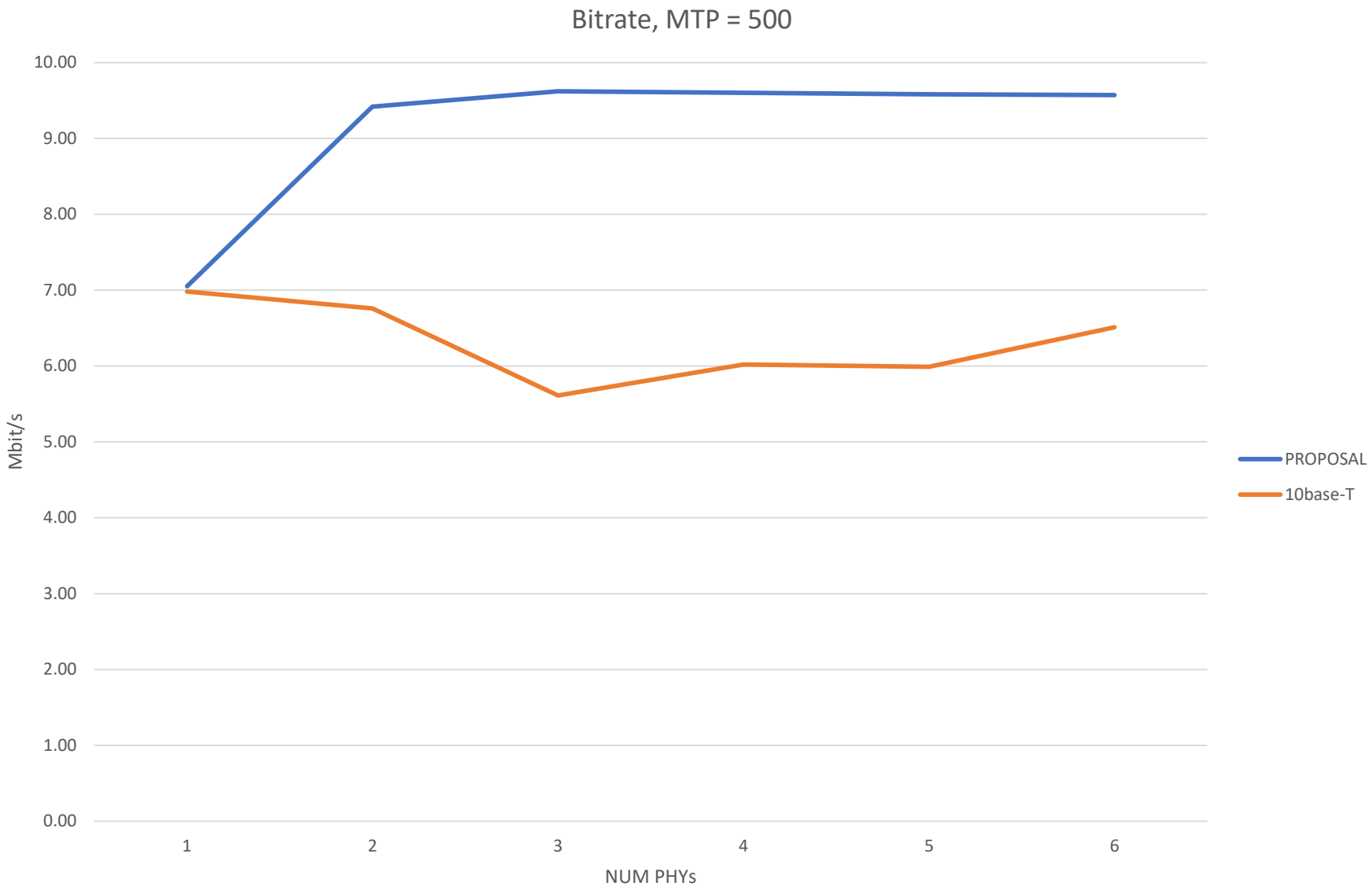
Simulations



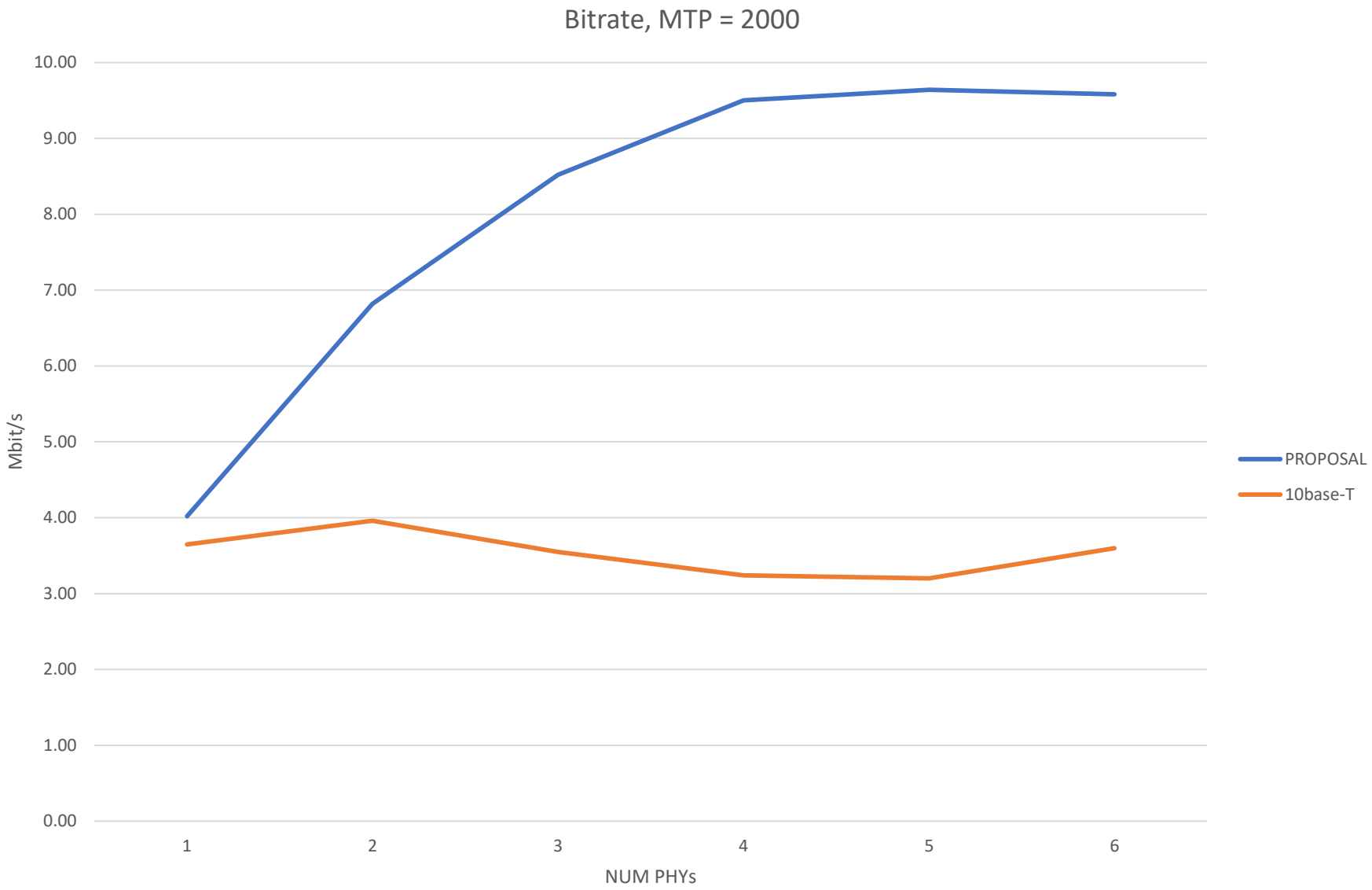
Simulations



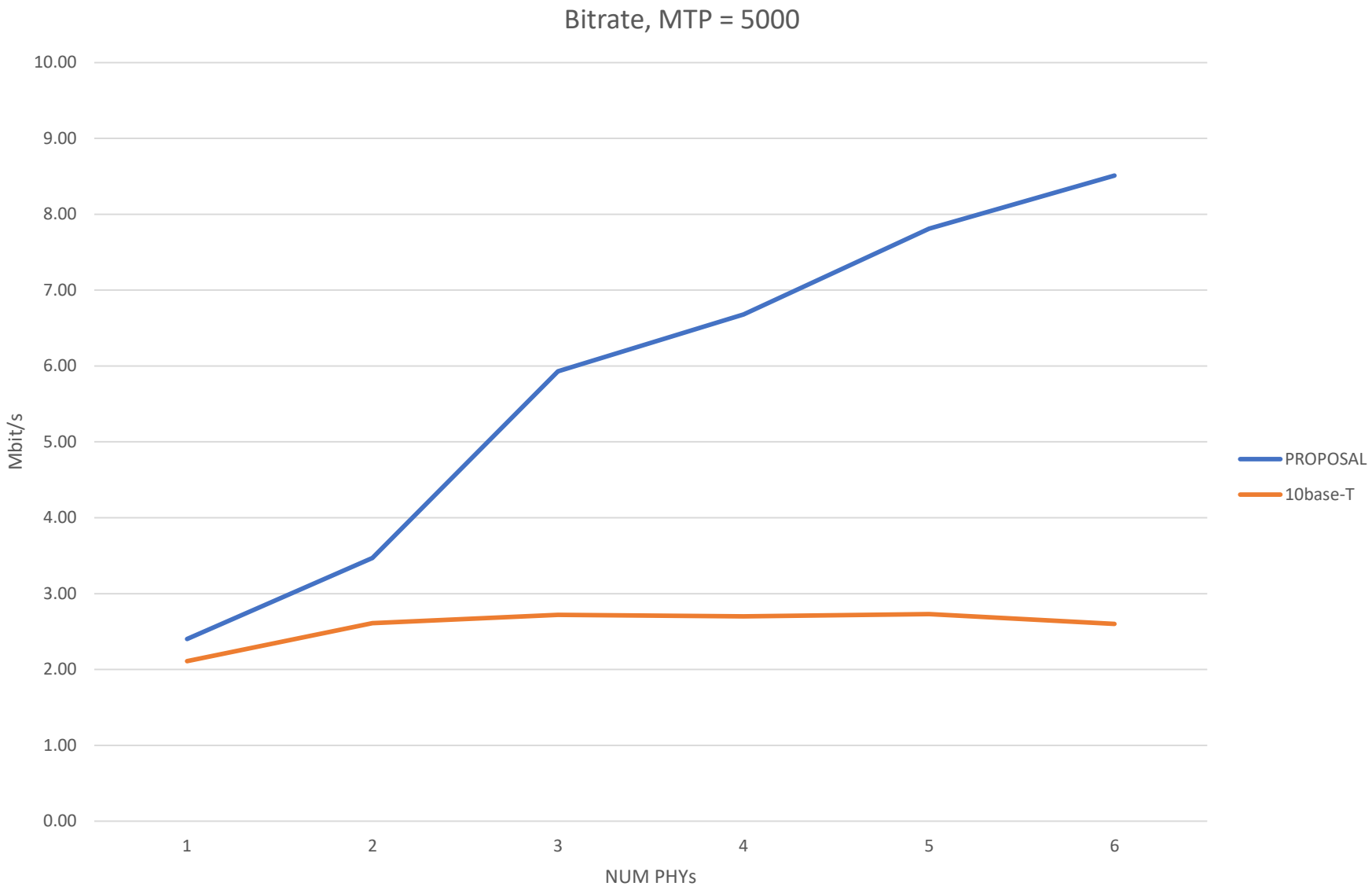
Simulations



Simulations



Simulations



Latency

- 500 pkts, size = 60B, variable MTP, 6 nodes. Latencies in µs.
- Comparison between simple half-duplex 10base-T and proposed PHY

MTP	MAX_LAT	AVG_LAT	STDEV
0	57595.6	553.3	4826.0
500	59692.8	1034.2	4637.4
2000	29387.5	618.9	2298.2
5000	19645.4	264.0	1035.7

10base-T

MTP	MAX_LAT	AVG_LAT	STDEV
0	443.4	441.1	26.2
500	546.4	186.4	90.7
2000	269.2	74.8	31.6
5000	223.7	64.0	17.8

This Proposal

Thank You !

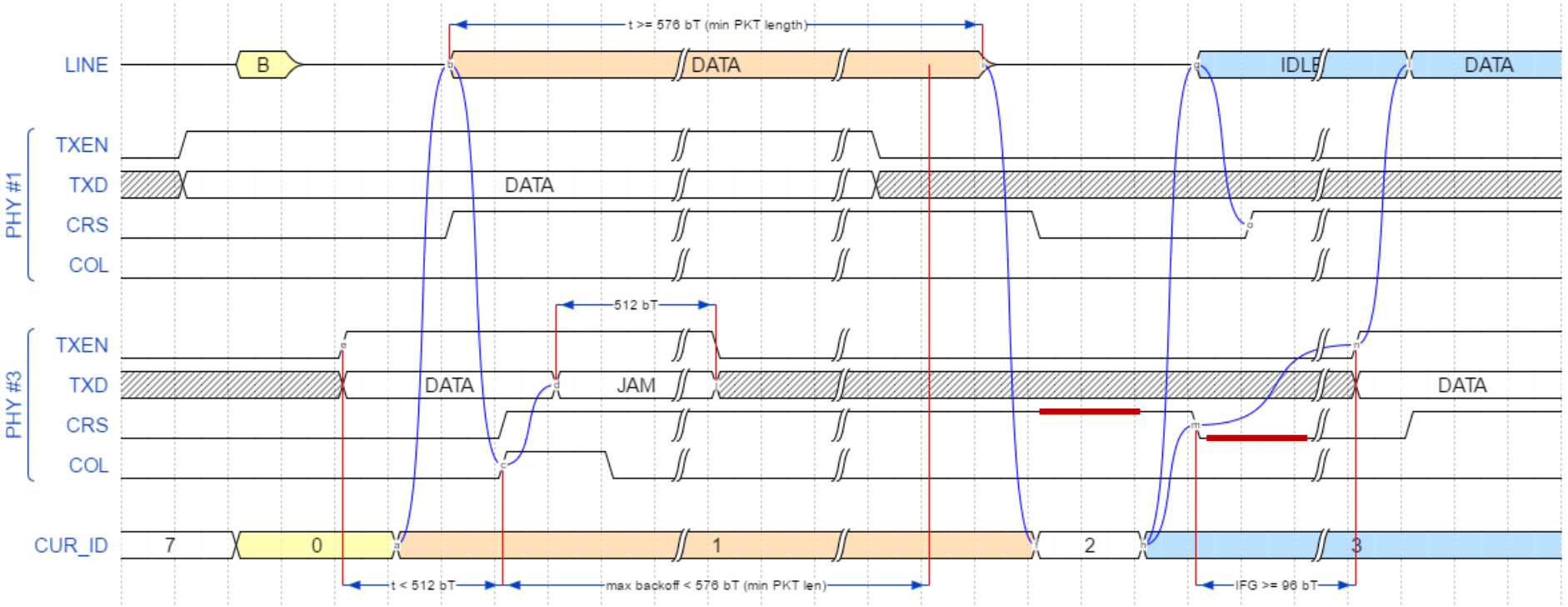
Interface with CSMA/CD MAC

- CSMA/CD MAC transmit process (from IEEE 802.3, clause 4)
  - If line is busy (CarrierSense = 1) → wait (defer transmission)
  - Wait IPG (at least 96 bits)
  - Start transmitting, despite line becoming busy again
    - If a collision is detected (COL = 1) → backoff:
      - Send jam for 32 bit times, stop transmission
      - Retry after random(0, ATTEMPTS) * 512 bit times
      - If ATTEMPTS > attemptLimit → give up (discard packet)
- CRS / COL can be used to have the MAC defer transmission until next handshaking
  - Use CRS to have the MAC defer transmission
  - Use COL at most once and only at beginning of a packet
    - MAC is ready to re-send in at most 32 + 512 = 544 bit times
      - Less than minimum packet size (576 bits)

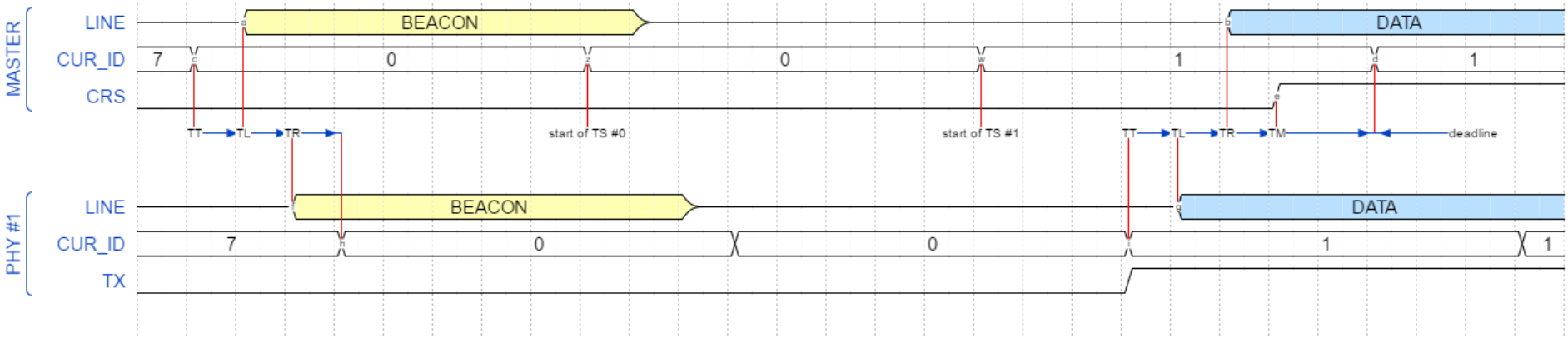
Join/ Leave

- Nodes in shut-down (e.g. link_control = FALSE) are silent
  - no special action required
  - good for energy efficiency
- Nodes joining the bus shall stay silent until first BEACON is received (see TX-PROCESS).
  - unique IDs are assumed to be provided by Layer Management
  - this proposal does not preclude auto-negotiation of the IDs, e.g. using LLDP

Example Waveforms



Clock Skew



PARAM	DESCRIPTION
TT	PHY TX latency
TL	Line propagation delay
TR	PHY RX latency
TM	margin

- Total clock skew = 2 * (TT + TL + TR)
- Margin > 0 for the system to work
- TS_TIMER > TOTAL SKEW