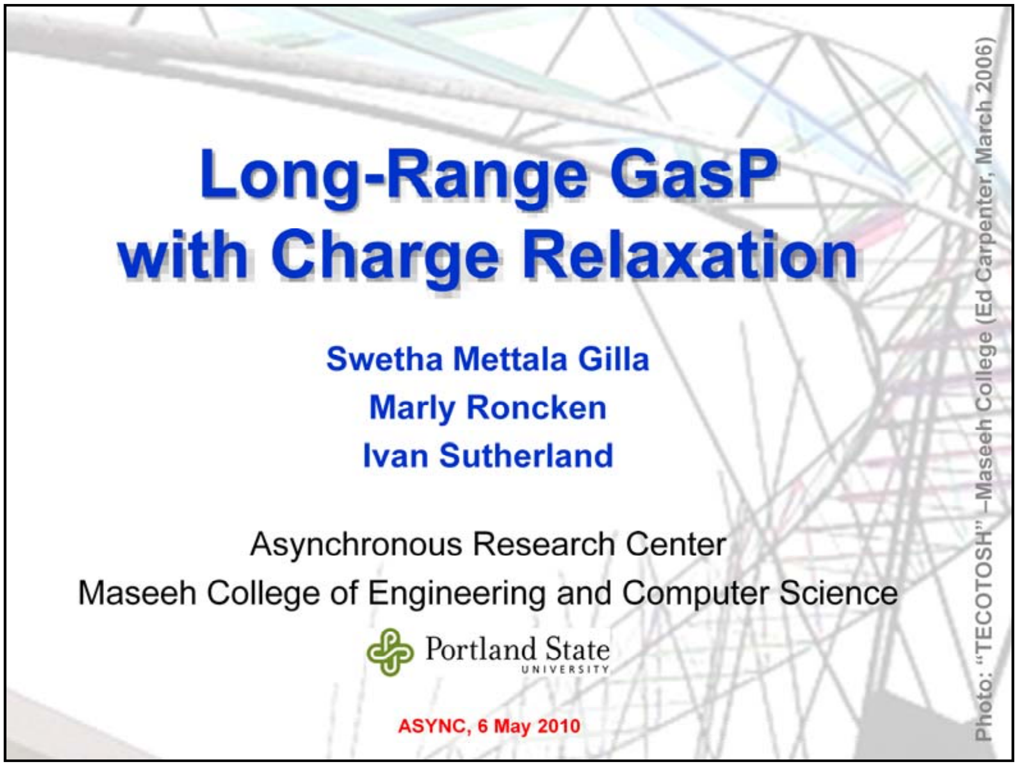


Good afternoon!
• My name is Swetha Mettala Gilla – you can call me Swetha.

• I'm a student at the Asynchronous Research Center at Portland State University,
• where I work on the timing of GasP modules.

I have been exploring how far apart we can place these modules, and still expect them to function.

• In this talk I will show how Charge Relaxation lets GasP modules function over long distances.
• I have done this work with Marly Roncken and Ivan Sutherland.

(PAUSE)
In the background of this slide, you can see faintly a sculpture.
(POINT)
It is called TECOTOSH.

Here are more pictures of TECOTOSH.
(PAUSE)

It stands in front of our engineering building at Portland State University.

• TECOTOSH explores MECHANICAL engineering limits
• and in a similar way - I explore ELECTRICAL engineering limits.

In this talk, I will explore the ELECTRICAL limits of single-track handshaking.

Introduction

2008 Infinity chip (90nm CMOS by TSMC):

- 10% degradation of peak throughput
- for 5000 versus 500 lambda long interconnect
  - *lambda = 50 nm*
  - *and so: 5000 lambda = 0.25 mm*

First wire delay study (Prasad Joshi – USC):

- Logical Effort + Lumped Capacitance model
- confirms measured observations
- inadequate for wires above 5000 lambda

ASYNC 2010, Long-Range GasP Slide 3

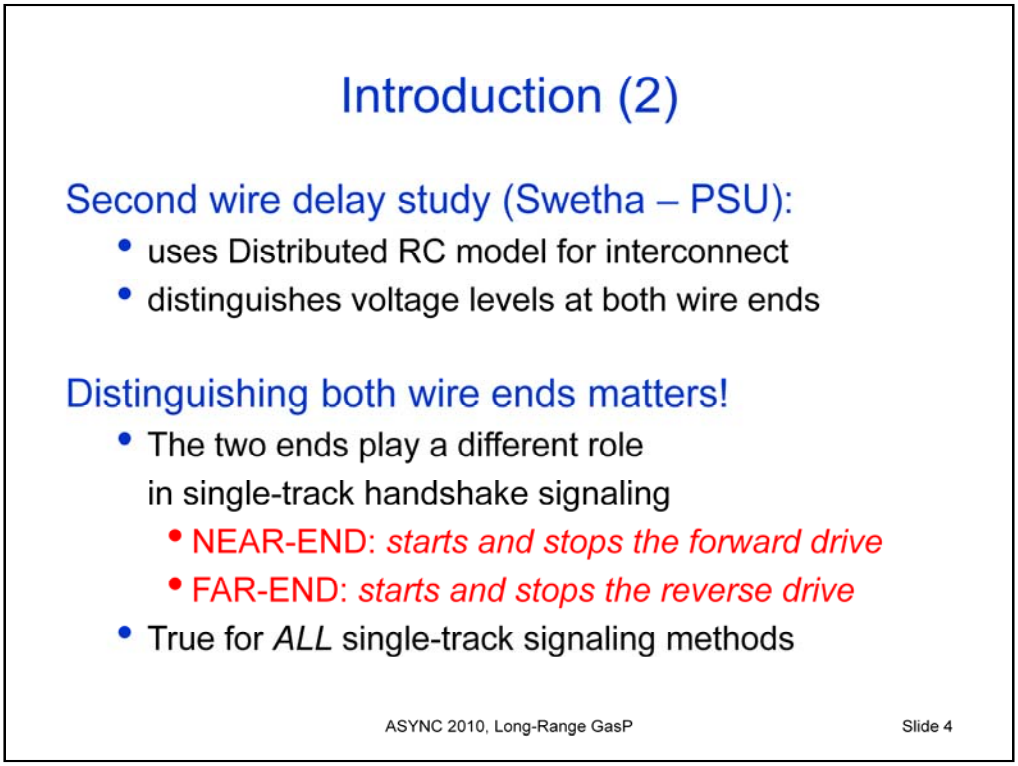
The motivation for this work comes from a 2008 test chip, called Infinity.
• The chip contains GasP experiments in 90nm CMOS, built by TSMC.
• One of these experiments has 10% less throughput than expected.

This was caused by a longer than usual single-track wire.
• It was 5000 lambda instead of 500 lambda.
• A wire length of 5000 lambda in this process is about a quarter millimeter.

To understand the theory behind this loss in throughput
• Marly and Ivan started a first wire delay study with Peter Beerel's group at USC.
• Peter's student, Prasad Joshi, combined the theory of Logical Effort with a Lumped Capacitance Model for wires
• Prasad's lumped capacitance model explained the chip measurements.

• The lumped capacitance model works for wires up to about 5000 lambda
• which is good enough for the Infinity chip.

But to obtain a deeper understanding, Marly and Ivan started a second wire delay study to look at longer wires.
• And this is where **I** came in.

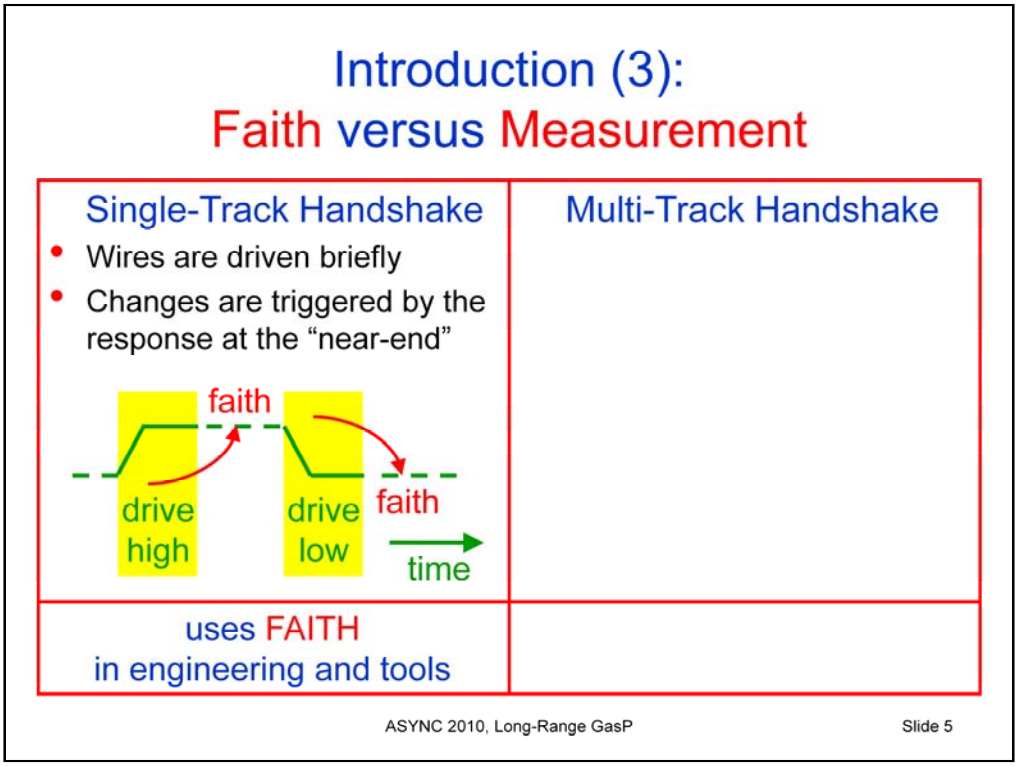


In this second wire delay study I use a distributed RC model for the single-track wires.

This is a more accurate model for longer wires because
• it models Resistance as well Capacitance
• and it can model different voltages at the two ends of the wire.

This matters in a single-track handshake, because the two ends play different roles:
• The voltage at the near end of the wire is used to start and stop the forward drive
• The voltage at the far end is used to start and stop the reverse drive.

• This is true for single-track communication in GasP
• AND it is true for other single-track communication methods as well.



To explain this further, here is a slide that compares single-track to multi-track handshaking.

Let's start with single-track.
(CLICK)
• A single-Track handshake uses the same wire for both directions.
• The wire has two drivers – one at each end.
• This leads naturally to a 2-phase return-to-zero handshake
(POINT TO the UP and DOWN transition in the PICTURE)

(WAIT)

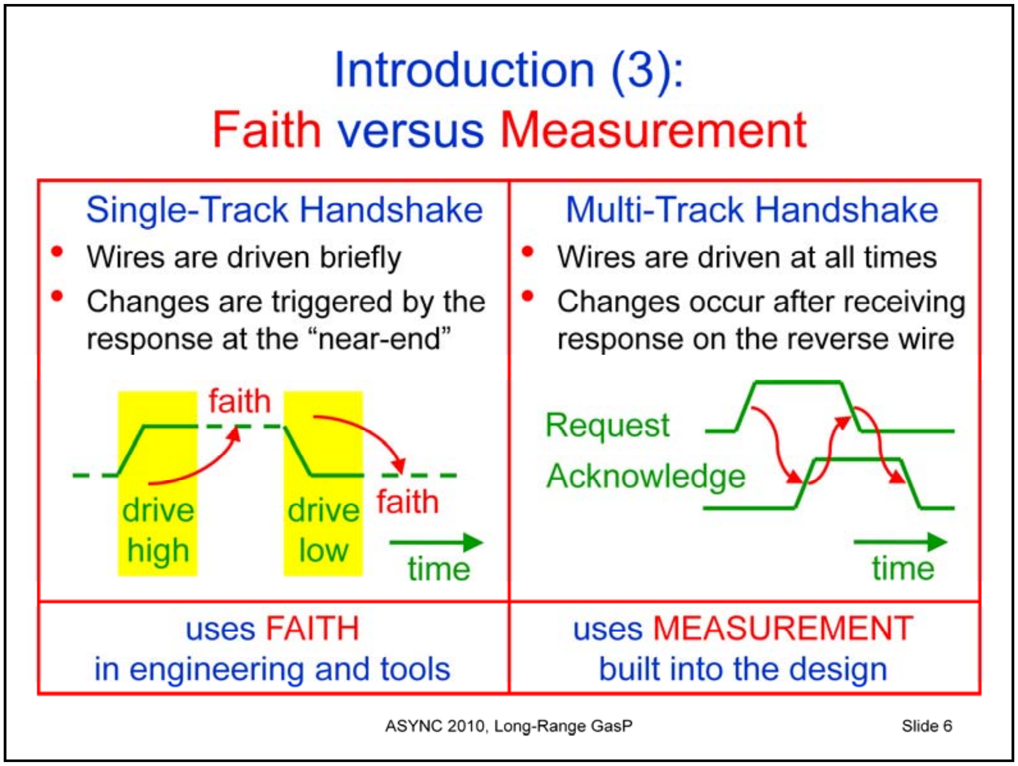
The wire is shared, so each driver must drive ONLY briefly,
• as indicated by the high-lighted areas in the picture.
• The brief drive avoids a drive fight between the two ends.

In addition, each driver responds only to changes at ITS OWN end.
• The local response completes the communication in two phases.

To make this work, we need some faith:
(CLICK)

Faith in our own engineering and in the design tools that we use.

Our faith in a single-track handshake depends on two assumptions:
• Assumption 1 is that the brief drive is long enough to traverse the wire.
• Assumption 2 is that the voltage observed at the NEAR-END of the wire reflects the voltage at the FAR-END.

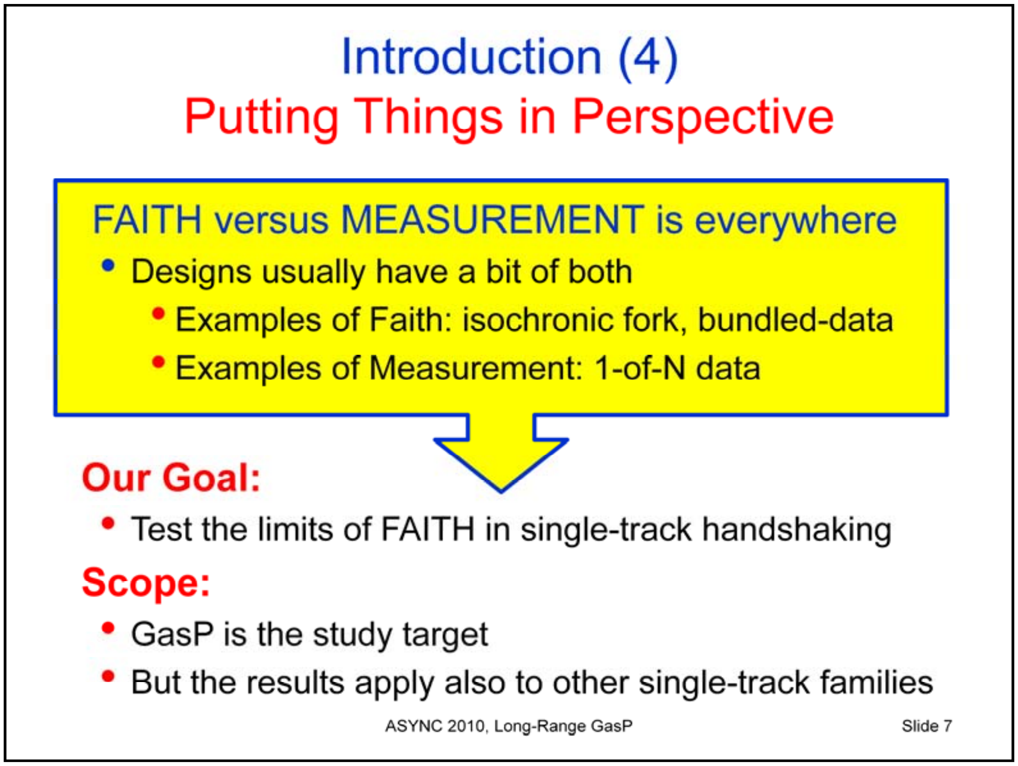


Compare this to multi-track handshaking that uses two separate wires.

Each end has one driver for one of the wires.
As a result:
• the wires are always driven, and
• changes always acknowledge what happened at the other end.

This leads naturally to a 4-phase handshake
• or a 2-phase handshake that doesn't return to zero.
(CLICK)

• Multi-track handshake signaling works by measurement built into the design.
• This offers good delay-insensitivity.

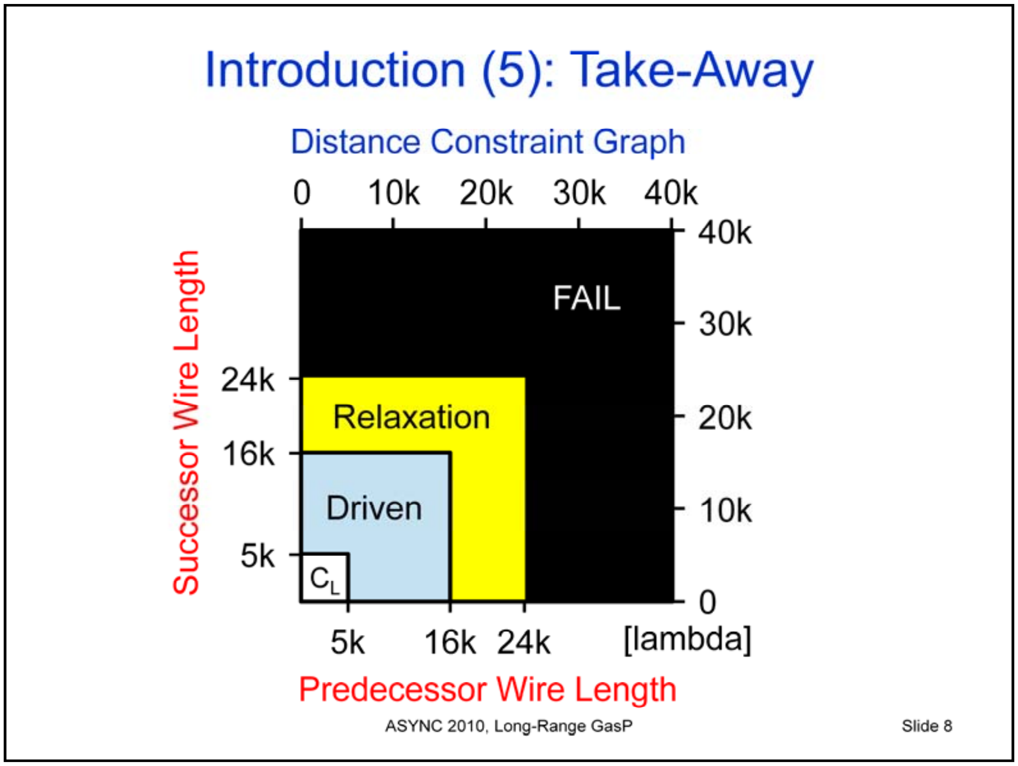


You may think that Faith and Measurement require radically different ways of designing.

But the reality is that designs usually have a bit of both:
• Think of isochronic forks, bundled-data, and the Relative Timing assumptions of a NAND-based C element.
• These are examples of FAITH.
• Faith in being able to place and route the design to satisfy such timing assumptions.

For high-speed GasP, we use FAITH where we CAN
and MEASUREMENT where we MUST.
(CLICK)

My study is about testing the limits of FAITH in single-track handshaking.
• GasP is my study target.
• But, my results apply also to other single-track families.



This is my last introduction slide, with the take-away for this talk.

We call this picture a “distance constraint graph”.
• It plots the length of the predecessor and successor wires from a GasP module.
(POINT to RED predecessor and successor axes)

• The colors show the short, medium and long wire regions
• Different wire models apply to these regions.

The surprising result of my study is that GasP works in THIS YELLOW region.
(POINT)
• Let me explain.

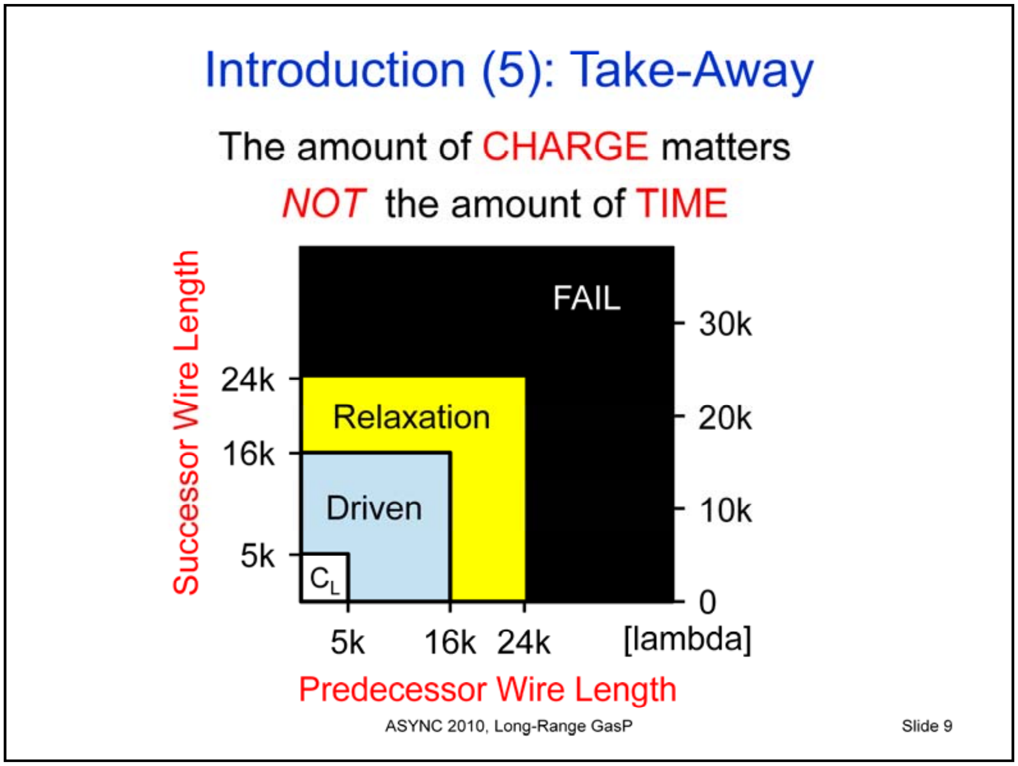
The lumped capacitance model is valid only for short wires
represented by the small white box.
(POINT)

The distributed RC model works better.
• In the blue region, the wires are shorter than 16000 lambda.
(POINT)

In this region, the RC model predicts that the wire delay is LESS THAN the drive time.
• As a result, the signal gets to the far-end before the near-end stops driving.
• The blue region is comfortable for Gasp and single-track handshake signaling.

The surprising result is for wires between 16000 and 24000 lambda – the yellow region.
(POINT)

The surprise is that GasP modules can operate at all with such long wires.
• In this region, the RC model predicts that the wire delay is GREATER THAN the drive time.
• As a result, the drive stops BEFORE the signal gets to the far-end.

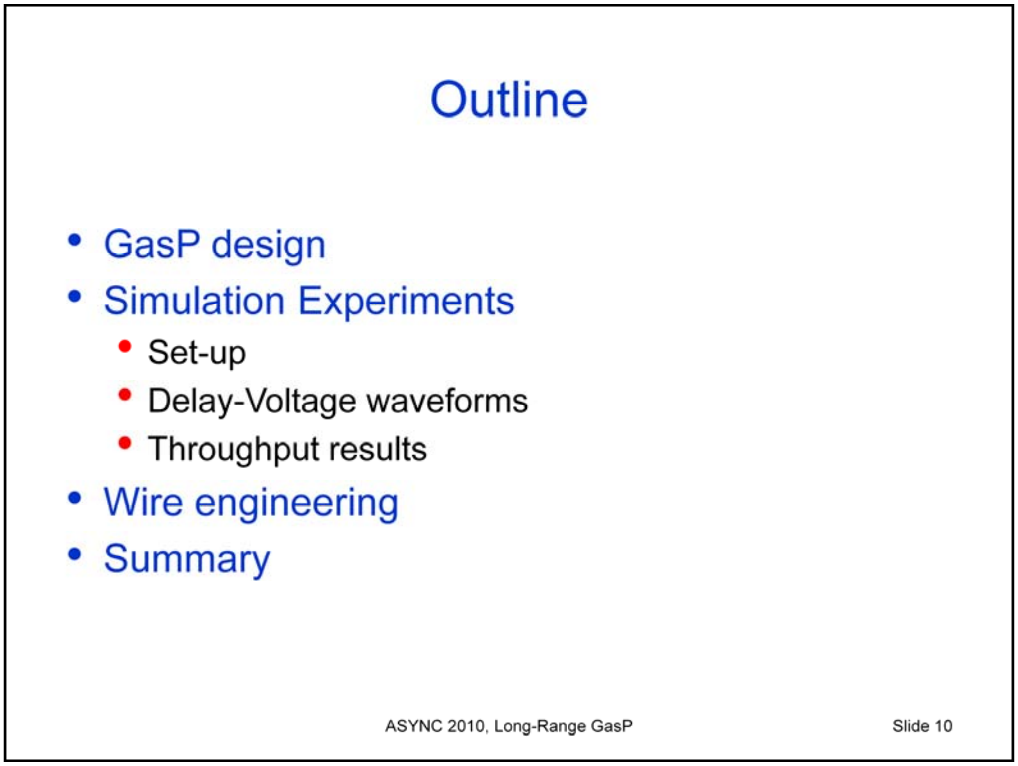


(NO CLICK needed: animation starts with slide)

This is OK because
- it's the amount of CHARGE that matters
- NOT the amount of TIME.

- Even after the drive stops, the charge in the wire spreads out over its entire length.
- This "charge relaxation" can deliver a signal without further drive.

Charge relaxation extends the range of GasP and single-track handshake signaling.
- That's the take-away.

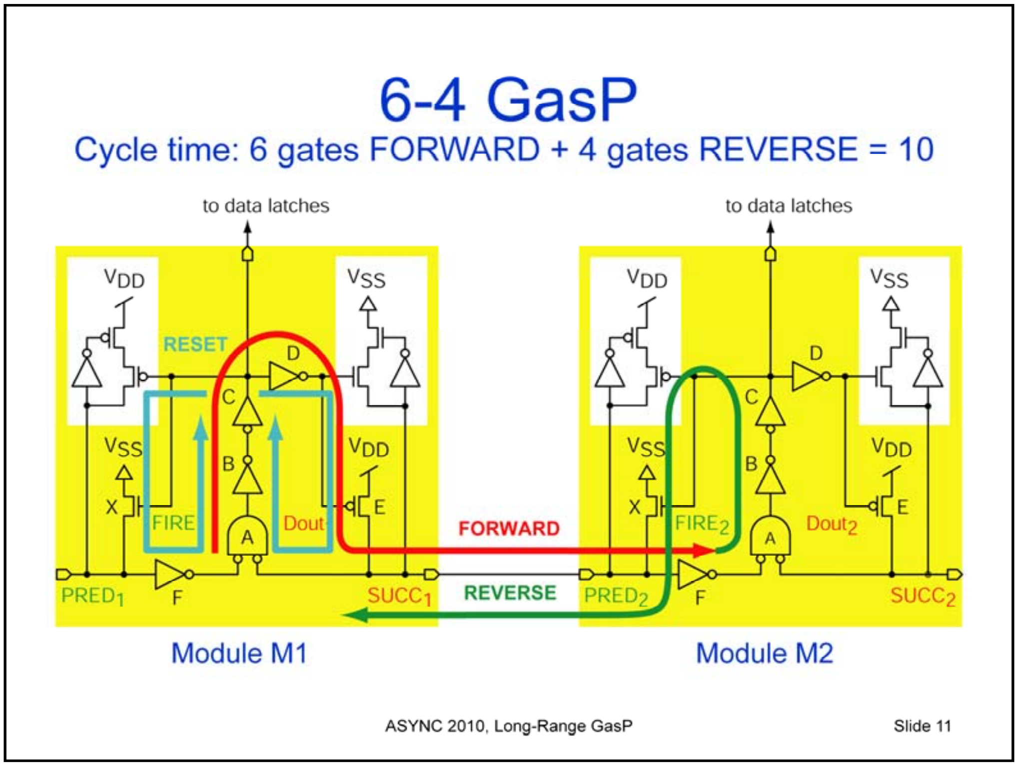


I have now told you
• What we did
• Why we did it
• and what we found.
The rest of my talk covers the simulation details.

• I will start with design details on GasP.

• After that I will show the simulation setup,
• I will go over the waveforms that we observed
• and I will show how longer wires limit throughput.

I will end with some comments on wire engineering and a summary.



We use 6-4 GasP.
• This slide shows two 6-4 GasP modules, M1 and M2, forming a 2-stage FIFO.
• The bundled datapath is omitted.
• The modules are connected by a single-track handshake wire in the middle. (POINT)
• Module M1 can drive the wire high via the P-transistor labeled E.
(POINT TO E)
• Module M2 can drive the wire low via the N-transistor labeled X.
(POINT TO X)

In GasP, we call the single-track wire a state wire, because it holds state.
• It is high when there is valid data - we call this full.
• And it is low when there is a bubble – we call this empty.

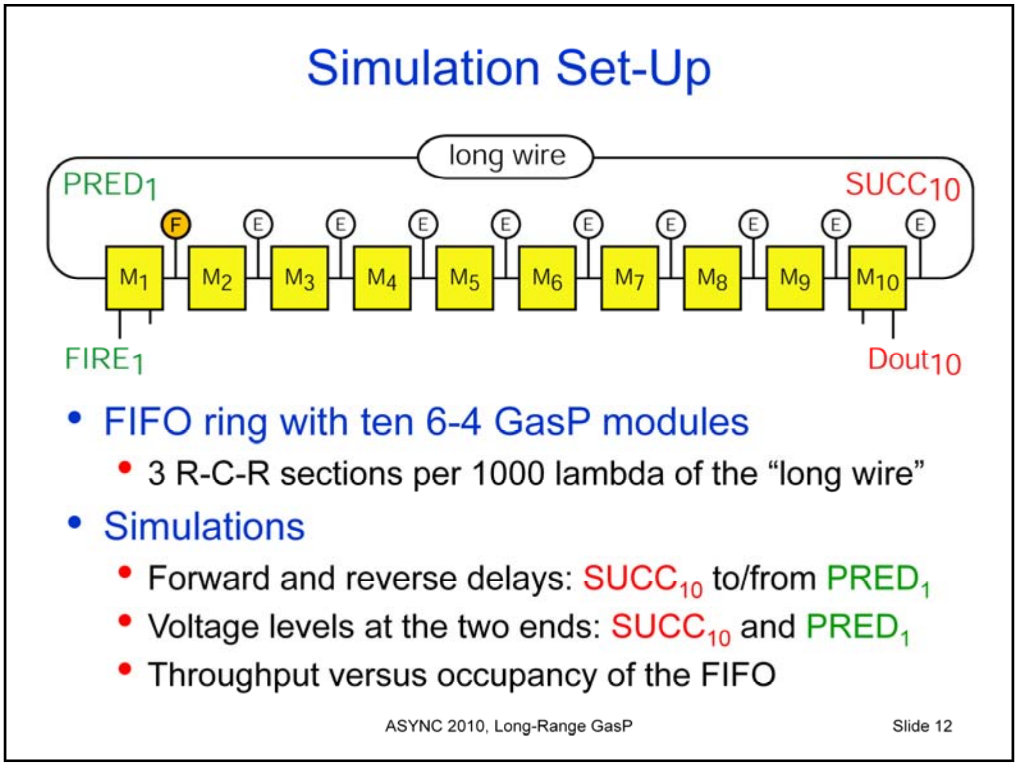
GasP module M1 acts when its predecessor is high, meaning full,
and its successor is low, meaning empty.
• When M1 acts, it forwards the full state to M2.
(CLICK)
• using the 6 gates on the red arrow: namely ABCDEF (POINT while naming)
• As soon as M2 can act, it drains the state wire.
(CLICK)
• using the 4 gates on the green arrow: namely ABCX. (POINT while naming)

The name 6-4 GasP comes from the 6 gates in the forward direction and the 4 gates in the reverse direction.
• The total cycle-time is 10 gate delays.
(CLICK)

In addition, each 6-4 GasP module has two self-resetting loops of 5 gates.
(CLICK)
• These two loops control the drive transistors E and X (POINT to left E and X)
• making sure that they drive only briefly.

When the single-track wire is not driven, its state is held by the half-keepers at the two ends of the wire,
• shown in these white areas (POINT TO the two MIDDLE KEEPERS)

(WAIT 4-5 seconds, look at slides, look at audience, then GO)



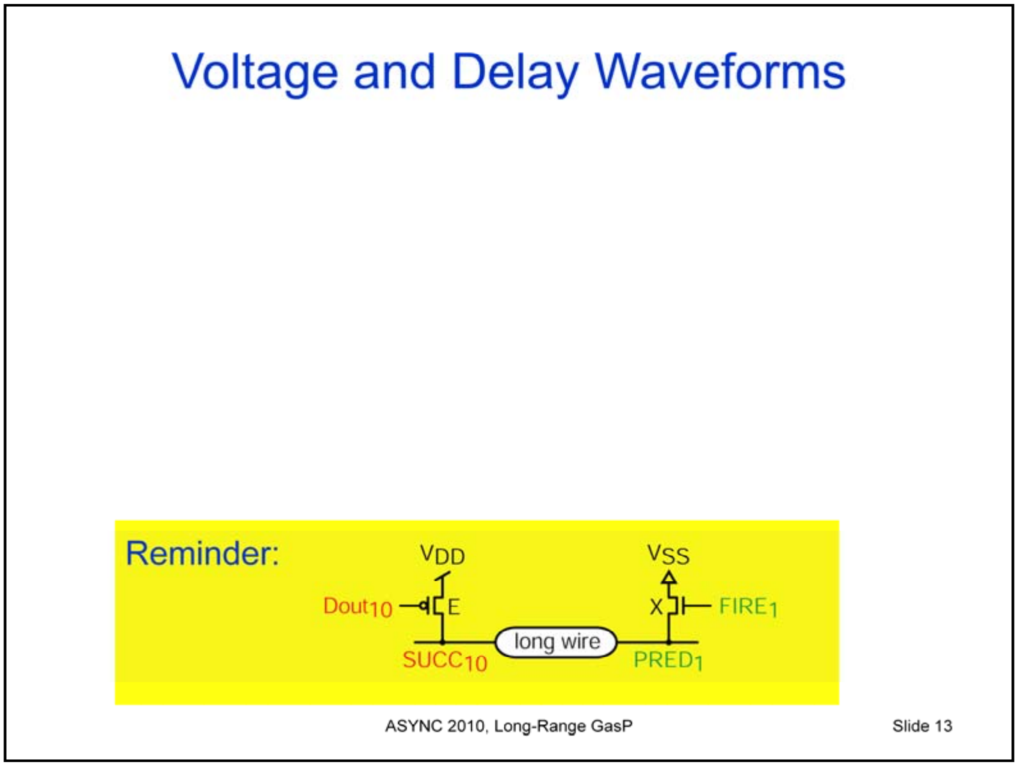
The setup for our simulations is a ten-stage ring FIFO.
• The yellow boxes M1 to M10 are ten GasP modules.
• You can see the state wires that connect them.

Each state wire has a little circle to set its START VALUE:
• The circles with E START the wire EMPTY, which is low.
(POINT to E-row)
• The circles with F START the wire FULL, which is high.

In this slide only the state wire between M1 and M2 starts out full.
(POINT to circle with F between M1-M2)
• And so, only one element will go around the ring.

Nine of the state wires are short – about 1000 lambda each
• The one at the top of the picture is much longer – and: that's the experiment.
(POINT to long wire)
• We model the long wire with an RC chain that has three R-C-R sections per 1000 lambda.

We use this setup to simulate three things:
• Forward and reverse delays through the long wire.
• Voltage levels at the two ends of the long wire.
• And a canopy diagram of RING-THROUGHPUT versus RING-OCCUPANCY.

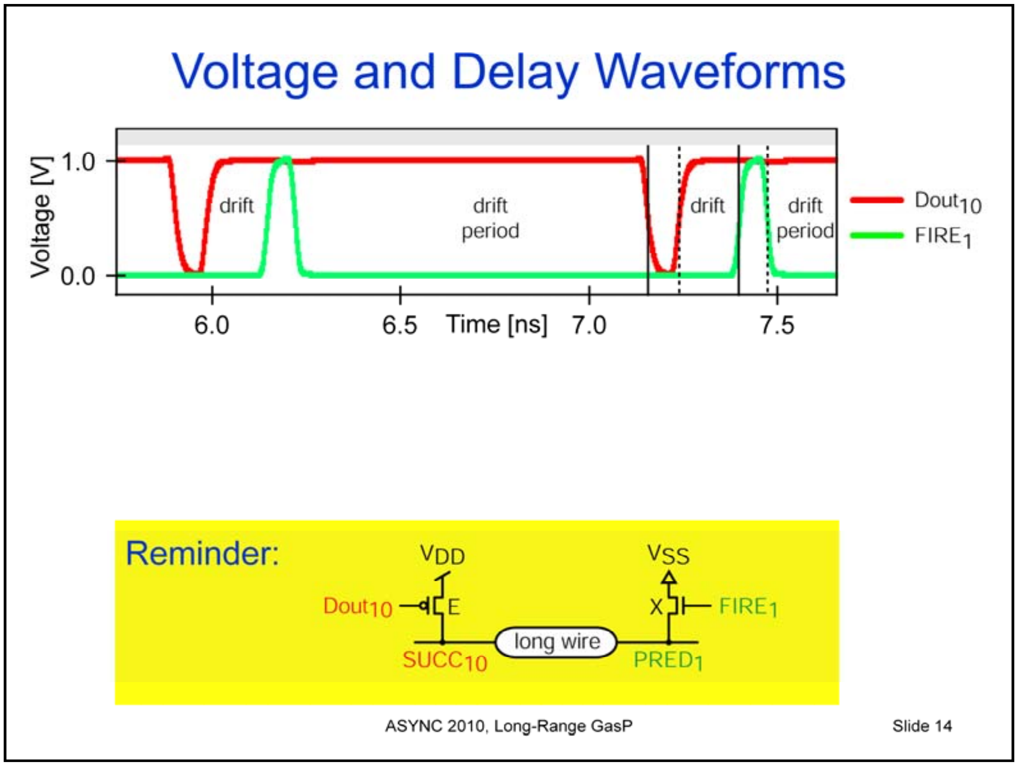


I will show you the waveforms for a wire that is 24000 lambda long.

The picture on the bottom of this slide shows the drive transistors for each end of the wire.
• I use RED for the FORWARD drive
• and GREEN for the REVERSE drive.

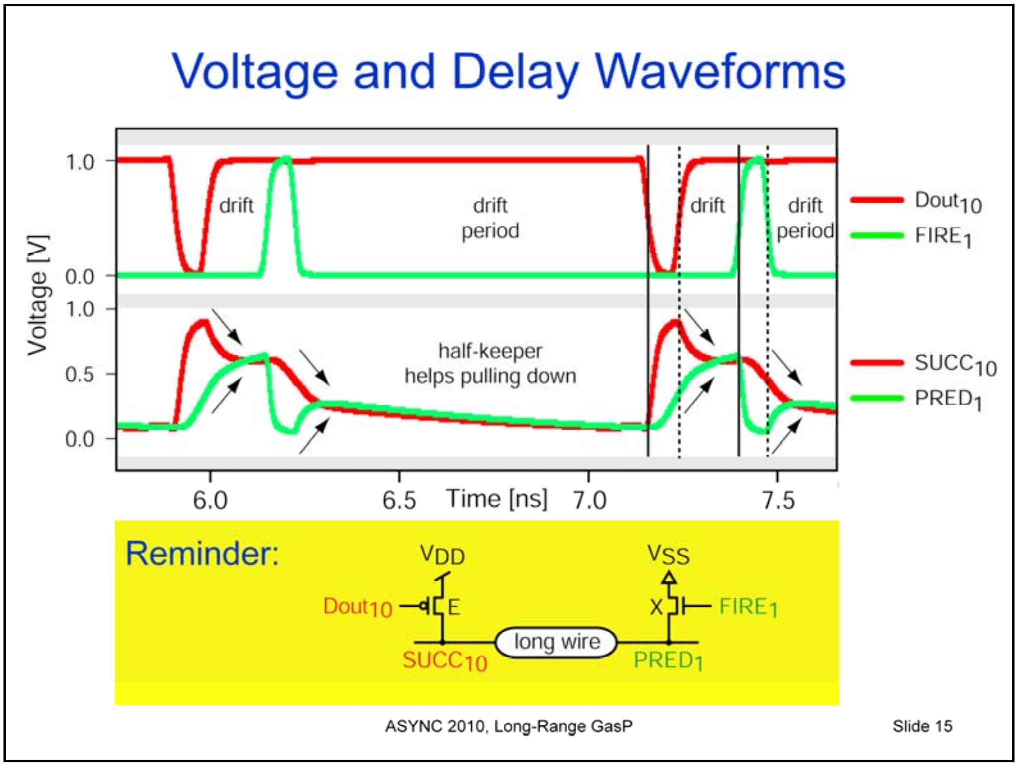
• Red-low turns on the P-transistor, driving the LONG WIRE high.
(POINT to red E transistor arrangement)
• Green-high turns on the N-transistor, driving the LONG WIRE low.
(POINT to green X transistor arrangement)

The colors correspond to the RED and GREEN waveforms on the next slide.



The waveforms at the top of this slide show the voltages to the drive transistors.
(POINT to top window)

- Notice that the top red signal is always above the top green signal.
- As a result, the two drive transistors never fight.



The bottom waveforms in the middle of this slide show the voltages at the two ends of the wire.
(POINT to middle picture with bottom waveforms)
• Notice that they are different.

Red shows the voltage at the output of the P-transistor.
• When the top red signal is low, the bottom red goes high.
(POINT VERTICAL top red to bottom red pulses)

(PAUSE)

Green shows the voltage at the output of the N-transistor.
• When the top green signal is high the bottom green goes low.
(POINT VERTICAL top green to bottom green pulses)

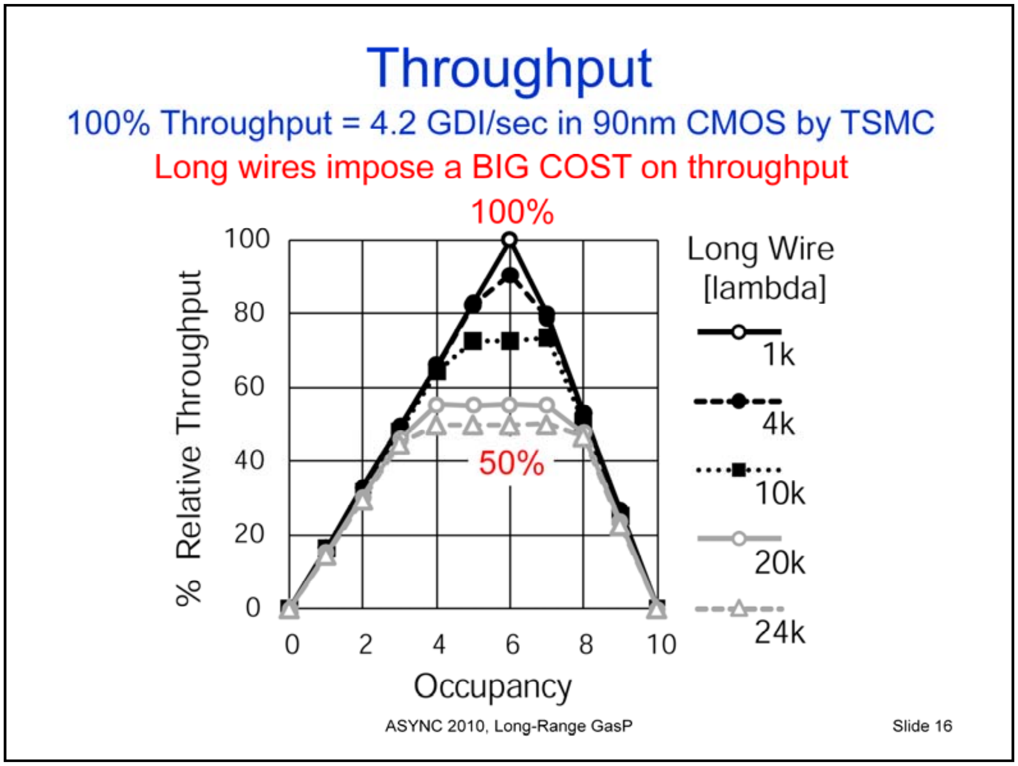
Between the top red and green drive pulses there is a period without drive, called DRIFT.
(PAUSE to POINT to first drift period)

• During the DRIFT period the charge in the wire relaxes throughout the wire.
• This charge relaxation brings the bottom red and green ends to nearly the same voltage
• as indicated by the arrows.
• HERE (POINT to 1st pair) and HERE (POINT to 2nd pair)
• and again HERE (POINT to 3rd pair) and HERE (POINT to 4th pair)

There is a longer drift period near the center of the picture.
(POINT vertical to long drift period in top and middle picture)

• For this period, you can see the action of the keepers.
• The keepers gradually pull the LONG WIRE to the intended voltage level
• which in this case is LOW.

(WAIT 4-5 seconds, look at slide, then at audience, then GO)



This is a simulated canopy diagram for the 10-stage GasP ring.

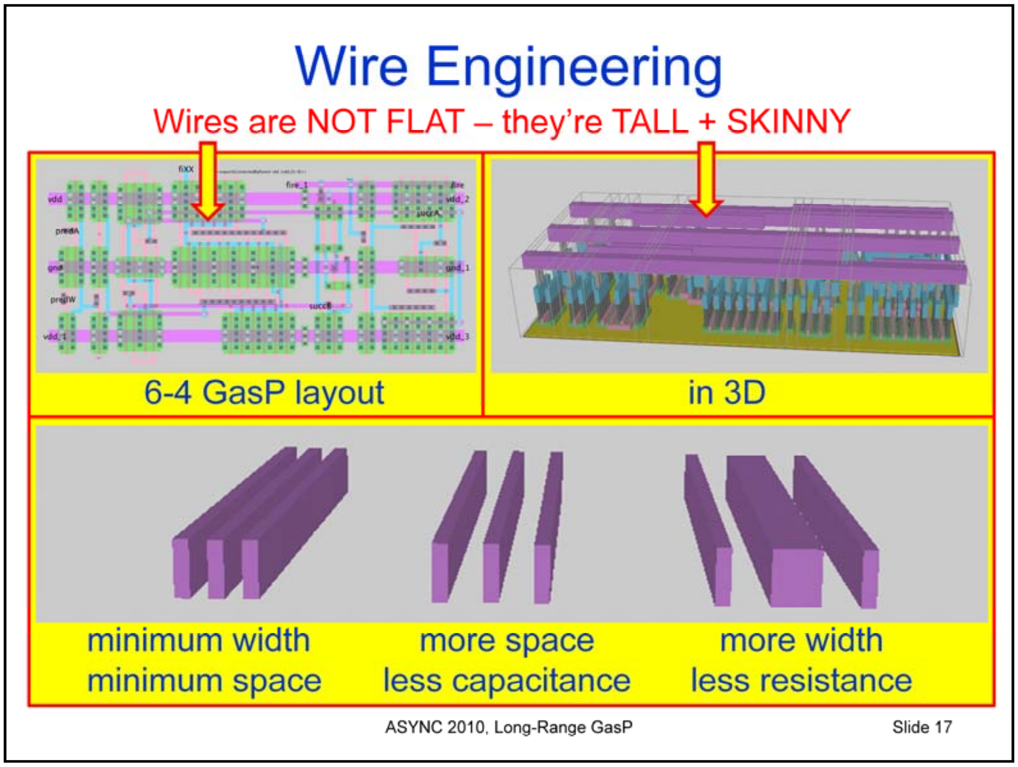
The top diagram with circles shows the throughput with the long state wire set to only 1000 lambda.
• From the Infinity chip
• we know that this throughput is 4.2 Giga Data Items per second in 90nm CMOS by TSMC.
• We have marked it here as 100% (one hundred percent) throughput.
(POINT to red 100%)

Longer state wires reduce the maximum throughput.
• The bottom diagram with triangles shows the throughput for 24000 lambda.
• The maximum throughput is now only 50%.
(POINT to red 50%)

(PAUSE)

SO, here's the PUNCH LINE:
• Although GasP will work with very long state wires, it is much slower.
(CLICK)
• Long wires impose a BIG COST on throughput !

(WAIT 4-5 seconds, look at slide, then at audience, then GO)



My results assume that the long state wire has minimum width.
(POINT to bottom left picture with min-width/min-space wires)
• We can do better.

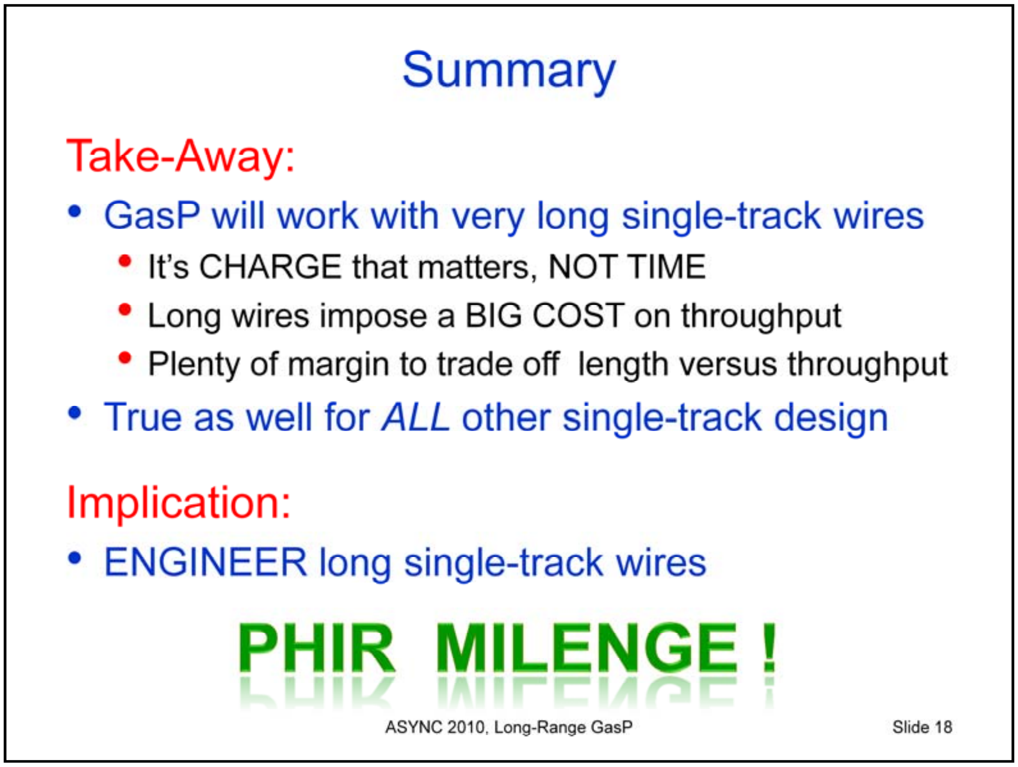
Take a look at the two pictures at the top of this slide.
• The left picture shows a layout of a 6-4 GasP module.
(POINT to left figure)
• The right picture shows the same module viewed in 3D.
(POINT to right figure)

Notice that wires are NOT FLAT as they seem in the layout.
(POINT to layout wire at the end of the left arrow)
• On the contrary: they are TALL and SKINNY.
(POINT to 3D wire at the end of the right arrow).

The consequence is that most of the capacitance is between adjacent wires.
• Therefore: if you separate a wire from its neighbors you can reduce its capacitance.
(POINT to bottom middle picture with more-space/less-cap wires)

Moreover: if you make the wire wider, you can reduce its resistance,
• without much change to its capacitance.
(POINT to bottom right picture with more-width/less-resistance wires)

For long-range GasP we believe that we "MUST ENGINEER" single-track wires.
• If we engineer only single-track wires but not bundled-data wires, the area overhead can be very small.



I have stretched single-track wires beyond comfortable design – ALL THE WAY to the point of failure.
• I have shown that GasP will work with very long single-track wires.
• AND that it's the amount of CHARGE that matters
• NOT the amount of TIME.

Although GasP will work with very long wires, it is much slower.
• Long wires impose a BIG COST on throughput !

(PAUSE)
When I showed these results to Ivan, Ivan's response was:
• No sensible designer would pay a 50% throughput penalty and risk HER design on charge relaxation.
• However:
• It is comforting to know that there is plenty of margin
• to trade off WIRE LENGTH against THROUGHPUT.
(PAUSE)

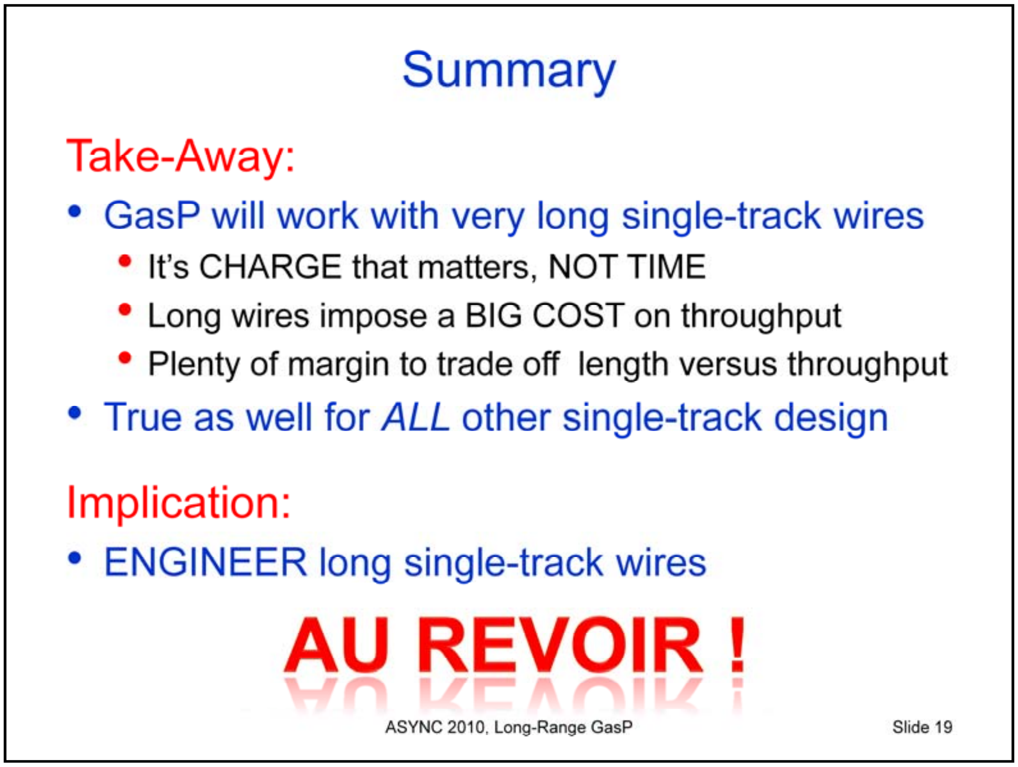
These results are true for single-track communication in GasP
• and they are true for other single-track communication methods as well.

(BIGGER PAUSE)

Perhaps the most important implication of my study is that
• we must engineer long single-track wires.

We're planning some experiments on our next chip to verify the results I've shown you today.
• We will also explore how wider state wires and greater separation can FURTHER extend the range of GasP.

I hope we will meet again to show you our next chip results.
• Until then, I'll say to you in Hindi:
(CLICK)
• "PHIR MILENGE"



Or as the French say:

• “AU REVOIR”
(pronounce as: “O REVWAR”)