

### VIVADO EXPERT SERIES | 2017



# High Speed Design Closure Techniques

Balachander Krishnamurthy

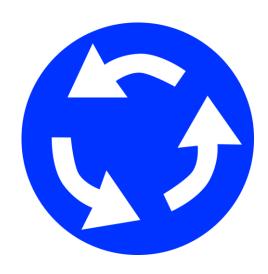


- High speed design challenges
- Design analysis
- Design guidelines
- Complexity and congestion analysis
- Summary

VIVADO EXPERT SERIES | 2017

## High Speed Design Challenges

- High performance requirement (>400 MHz)
- Issues detected too late
- RTL coding style
- Constraints not clean
- Too many logic levels
- Too many high fanout nets and too many control sets
- Sub-optimal clocking and big blocks (BRAM, DSP)
  - Trade-off for power vs performance
- Complexity and Congestion
- Many timing closure iterations



# UltraFast Design Methodology (UG949)

- WHAT WHERE WHY HOW ANSWERS
- Iterate after synthesis for faster analysis and predictable results
  - RTL Coding style and examples
- Recommendation with checklists in all areas:
  - Board and device planning
  - Design creation
  - Implementation
  - Design Closure
- Run Report Methodology to identify and fix issues early
  - Synthesis, Timing, XDC and Clocking methodology checks



## UG1231: UltraFast Design Methodology Quick Reference Guide

#### UltraFast Design Methodology Quick Reference Guide (UG1231)

#### INTRODUCTION

The UltraFast™ Design Methodology is a set of best practices recommended by Xilinx to maximize productivity and reduce design iterations of complex systems, including embedded processor subsystems, analog and digital processing, high-speed connectivity, and network processing. See the UltraFast Design Methodology Guide for the Vivado Design Suite (UG949) for more information.

The UltraFast Design Methodology Checklist (XTP301) includes common questions that highlight typical areas where design decisions have downstream consequences and draws attention to potential problems that are often unknown or ignored. It provides easy access to related collateral. The checklist is available within the Xilinx Documentation Navigator tool (DocNav).

This quick reference guide highlights key design methodology steps to achieve quicker system integration and design implementation and to derive the greatest value from Xilinx® devices and tools. Pointers to related collateral are also provided. The main design tasks covered in this guide include:

- Board and Device Planning
- Design Entry and Implementation
- Top-Level Design Validation
- Design Analysis
- Design Closure

Refer to the UltraFast Design Methodology – System-Level Design Flow available within the Xilinx Documentation Navigator tool (DocNav) for pointers to all design hubs and specific collateral.



UG1231 (v2016.3) November 11, 2016

#### BOARD AND DEVICE PLANNING

#### **PCB Designer**

#### Examine Key Interfaces

 Validate part orientation and key interfaces

#### Examine the PCB Layout

- Perform the Memory Interface and Transceiver Checklists
- Follow PCB layout recommendations
- Ensure final FPGA pinout is signed off by FPGA designer

#### Review the Schematic

- Complete PCB Checklist
  review
- Check PDS, configuration, and power supplies
- Validate I/O state before, during, and after configuration

#### Manufacture and Test

 Verify the configuration sequence, power supplies, and I/O performance with the test I/O project

#### See Also:

UG949: Board and Device Planning
PCB Design Checklist
Memory Interface IP Design
Checklists
Schematic Design Checklists

#### **FPGA Designer**

#### Analyze Device for Pinout

- Examine transceiver and bonded I/O locations
- Examine SSI technology I/O planning
- Validate part orientation and key interfaces

#### Define I/O Pinouts for Key Interfaces

- Create I/O planning projects
   Define and validate memory
- Define and validate memory controllers, GTs, and PCle\* technology locations
- Establish a clocking skeleton
- Minimize floorplan distance between connected IP

#### Define Final Pinout

- Merge interface projects into a final I/O project
- Validate DRCs and SSN analysis
- Implement design to check clocking and I/O rules
- Use the final I/O project for production test

#### Estimate Power

- Determine power budget and thermal margin using Xilinx Power Estimator (XPE)
- Apply toggle rates using knowledge of prior designs

#### See Also:

UG949: Board and Device Planning Power Estimation and Optimization Design Hub I/O and Clock Planning Design Hub

#### DESIGN ENTRY AND IMPLEMENTATION

#### Logic Designer

#### Define a Good Design Hierarchy

- Define relevant hierarchies to help global placement and floorplanning
- Insert I/O and clock components near the top level
- Add registers at main hierarchical boundaries
- Generate IP and review target device utilization

#### **Build and Validate RTL Submodules**

- Ensure design adheres to RTL coding guidelines
- Add sufficient registers around DSP and memories
- Use control signals only when absolutely necessary
- Use synthesis attributes to control final logic mapping
   Control simple timing and tributes to control final logic mapping
- Create simple timing constraints to review estimated timing and address paths with too many logic levels
- Review synthesis log files, utilization report, and elaborated view to identify sub-optimal mapping
- Run Methodology and RTL checks and review issues
- Implement the submodule in out-of-context (OOC) mode to validate implemented performance
- Review utilization and power against original budget
- Simulate the design to validate functionality

#### Assemble and Validate Top-Level Design

- Synthesize the top-level RTL design and resolve all connectivity issues
- Review top-level utilization and clocking guidelines
- Create and validate top-level constraints
- Iterate the RTL and constraints to fix Methodology and DRC issues and meet timing
- Proceed to implementation

#### See Also:

UG949: Design Creation and Implementation
Designing with IP Design Hub
Using IP Integrator Design Hub
Logic Synthesis Design Hub
Applying Design Constraints Design Hub
Implementation Design Hub



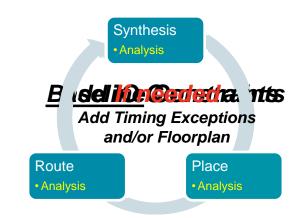
# Design Analysis

Constraints, logic levels, control sets and high fanout nets

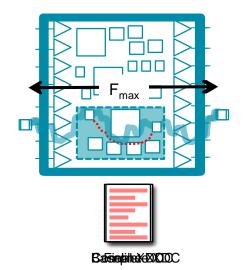
## Baselining – Clean Constraints

### Baselining: Ensure timing constraints are clean

- All clocks and clock relationships are defined
- Asynchronous clock domain crossings are constrained correctly
  - ➤ 1-bit CDC's: false path or clock group constraints
  - ➤ Multi-bit CDCs: Use set\_max\_delay -datapath\_only and/or set\_bus\_skew
- IO constraints
- Other point-to-point exception constraints



#### Optionina Eighte Electrical Planips





## Logic Levels

- Review number of logic levels
  - Guideline: 500 ps per logic level (1 LUT + 1 net)
  - Analyze with:

```
report_design_analysis -logic_level_distribution -extend
```

- Recommendations:
  - Modify RTL to reduce the number of logic levels
  - Use instance based synthesis to optimize modules
    - ➤ Synthesis Options: retiming, area, performance, routability, etc.

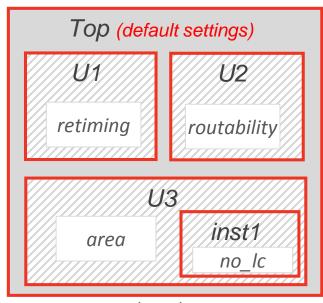
Speedgrade	-1	-2	-3
7 Series	575ps	500ps	425ps
UltraScale	490ps	425ps	360ps
UltraScale+	350ps	300ps	250ps
UltraScale+ LV	490ps	425ps	360ps

## Instance Based Synthesis Options

Instance-based optimization using XDC

```
set_property BLOCK_SYNTH.RETIMING 1
set_property BLOCK_SYNTH.STRATEGY {ALTERNATE_ROUTABILITY} [get_cells U2]
set_property BLOCK_SYNTH.STRATEGY {AREA_OPTIMIZED}
set_property BLOCK_SYNTH.LUT_COMBINING
```

- Nested options supported
- Strategies are presets
- Multiple logic options allowed



**Instance Based Synthesis Settings** 

Default Area Performance Routability retiming adder threshold comparator\_threshold shreg\_min\_size fsm\_extraction lut combining flatten inside partition extract partition flatten\_hierarchy control\_set\_threshold max\_lut\_input muxf\_mapping keep equivalent register

*Strategies* 

(bundled preset)

Logic Options

[get\_cells U1]

[get\_cells U3]

[get\_cells U3/inst1]

Strategies and logic options applicable to instances

### Control Sets

Analyze with report\_utilization and report\_control\_sets -verbose

### Control sets can impact placement and timing

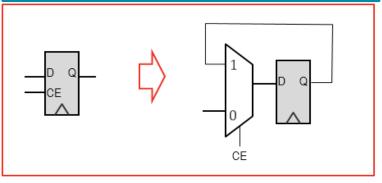
### Guideline

- Acceptable if unique control sets < 7.5% of total slices</li>
- Analysis needed if unique control sets > 15% total slices

### Recommendations

- Manually replicate registers with high fanout or PhysOpt
- Global synthesis option to change defaults (-control\_set\_opt\_threshold)
  - > Different thresholds based on architecture
    - √ 4 for 7-series and 2 for UltraScale and UltraScale+
- Implement timing critical low fan-out (<8) control signals in fabric</li>
  - By using extract\_enable or extract\_reset in RTL
  - > In XDC file:

```
set_property extract_enable/reset "no" [get_cells U1/tmp_reg]
```



CE emulation with logic on D input side

## **High Fanout Nets**

Analyze by running report\_high\_fanout\_nets

### RTL Recommendations:

- If timing is not critical and high fanout nets feed FFs directly
  - > Promote nets with fanout >25k to global clocking

```
set_property CLOCK_BUFFER_TYPE BUFG [get_nets netName]
```

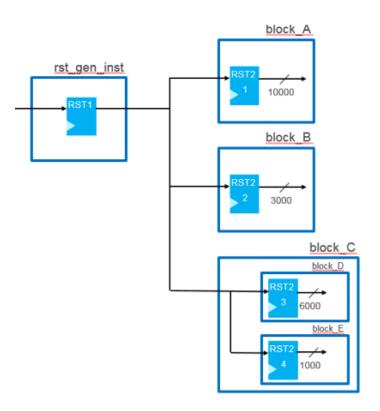
- Timing is Critical
  - ➤ Replicate registers in RTL with KEEP attribute
  - ➤ Replicate in each SLR with pblock constraints
- Avoid high fanout from/to DSP/BRAM

### Synthesis Recommendations:

- Avoid MAX\_FANOUT for global control signals
- Use MAX\_FANOUT only inside small module if needed

## Replicate high fanout net drivers with PhysOpt

- Phys\_opt\_design -force\_replication\_on\_nets
- Phys\_opt\_design -directive AggresiveExplore



## High Fanout Nets Optimization Flow

### Synthesis: Limited Replication

- Avoid MAX\_FANOUT for global control signals
- Use MAX\_FANOUT only inside small module if needed
- Use KEEP on manually replicated cells

## Opt Design: Coarse-grained Replication

Based on large design hierarchies

## Place Design: Mid-grained Replication

Based on early placement and timing information

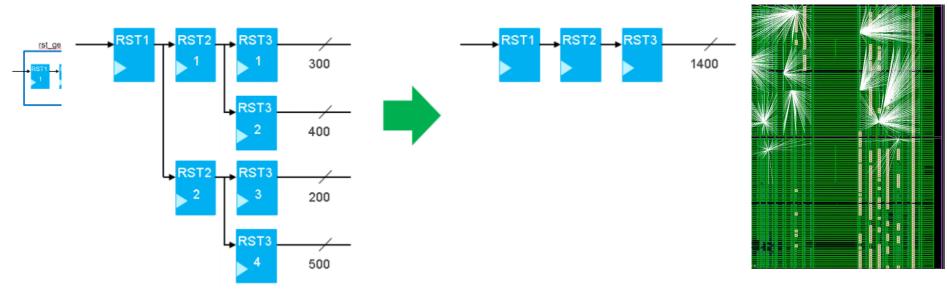
## Phys Opt Design: Fine-grained Replication

Based on accurate timing information

## High Fanout Nets Optimization in Opt Design

### Opt Design

- Insert, merge and split BUFGs on high fanout nets (Done by default)
- Merge low fanout control signals in opt\_design with '-control\_set\_merge'
- Merge all LUT and Flop equivalent drivers with '-merge\_equivalent\_drivers'
- Control module based replication with '-hier\_fanout\_limt <number>'



## High Fanout Nets Optimization in Placer

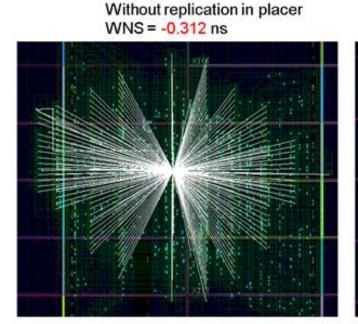
- Physical Synthesis in Placer (PSIP)
  - Automatically inserts BUFGs during global placement based on resource availability
  - Replicates high fanout nets during placement with '-fanout\_opt'
  - Replicates low/medium fanout nets connected to DSP/BRAM

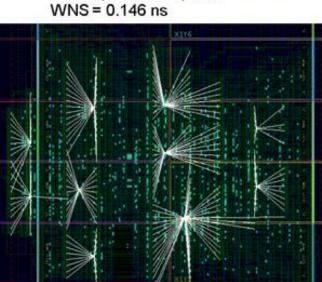
### Advantages

- No guess needed work at RTL/Synthesis level
- Optimal control set usage
- Placement is known and replication is based on driver and load placement
- Addresses timing critical HFN early in placer
- Reduces need for post placement physical optimization

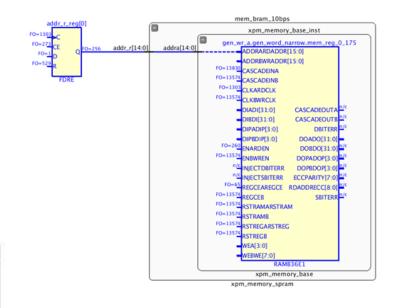
## PSIP Example : Control Signal

FF driving address bits of large BRAM





With replication in placer



# Design Guidelines

Clocking, Synchronous CDCs, BRAM and DSPs

## **Clocking Guidelines**

- Sub-optimal clocking hurts timing
  - Review and fix clocking methodology checks
- Reduce clock buffer utilization, where possible
  - Reduce clocks by combining identical synchronous clocks
  - Remove buffer from MMCM feedback path (no delay compensation support)
  - Remove cascaded buffers unless absolutely required
  - Remove buffer from MMCM feedback path when in INTERNAL mode
- Floorplan clocks to keep
  - Clock with low fanout close to the driver
  - Away from highly utilized areas
- Avoid clocks routed on fabric resources
  - Unpredictable skew, competes with routing resources with other nets
  - Insert clock buffers, if possible (loads >30)



## Guidelines for synchronous CDC

### Analyze critical crossings

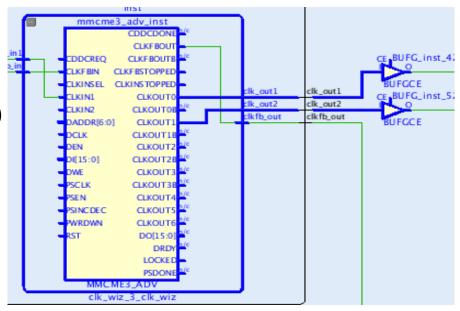
- Run report\_timing\_summary or report\_clock\_interaction
- Tight setup requirements (more sensitive to skew)
- Highest number of paths (largest impact on TNS & THS)

## Review and improve CDC clock topologies

- Configure VCO of MMCM to run at highest frequency
  - Reduces clock uncertainty
- Avoid common node before MMCM
- Avoid cascaded buffers
- Avoid crossing IO/SLR boundaries to minimize skew

## Reduce/relax CDC paths

- Treat as asynchronous (requires proper synchronization circuit)
- Use multicycle path constraints (relax setup) (requires CE logic)



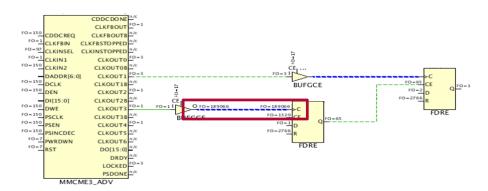
## Clock Skew in Synchronous CDCs

CLOCK\_DELAY\_GROUP constraint to limit skew on synchronous clocks

### Match clock net delays in a group

- Useful for constraining synchronous CDC skew
- Better accuracy than using USER\_CLOCK\_ROOT constraint

### • Example:



set\_property CLOCK\_DELAY\_GROUP group\_0 [get\_nets {u1/clk\_50M u1/clk\_100M}]

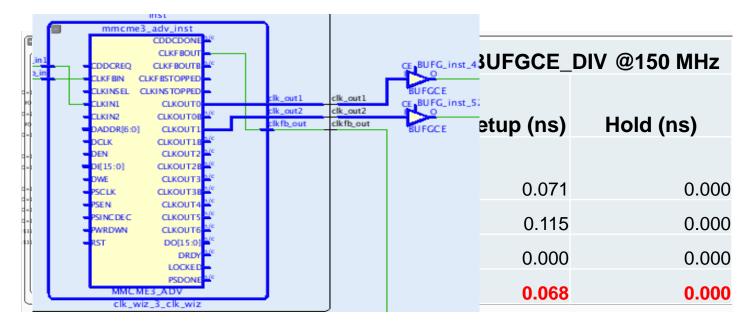
### Guidelines

- Apply to net segment directly connected to buffer output
- Buffers must have same driving cell => ensures balanced topology
  - ➤ Driver examples: MMCM, PLL, IBUFDS, GT CHANNEL



## Leverage BUFGCE\_DIV in UltraScale devices

- Use BUFGCE\_DIV to perform division
  - Removes phase error (~120ps) => helps both setup and hold on CDC paths
  - Clocks with simple period ratio (/1 /2 /4 /8)
  - Only 4 BUFGCE\_DIV per clock region
- Beware of cell delay difference between BUFGCE and BUFGCE\_DIV
  - Best if both clocks use same buffer type (BUFGCE\_DIV can divide by 1)

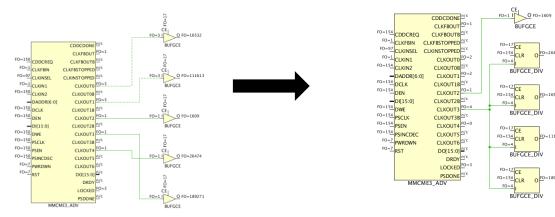


## Real customer design - Case Study

- Design with high speed clocks and many synchronous CDCs
- Identify critical CDC paths with report\_clock\_interaction (pre-route)

Source Clock	Destination Clock	WNS (ns)	TNS (ns)	Failing Endpoints (TNS)	Total Endpoints (TNS)	Path Req (WNS)	WHS (ns)	THS _ 1 (ns)	Failing Endpoints (THS)	Total Endpoints (THS)	Path Req (WHS)	Common Primary Clock	Inter-Clock Constraints
svs clk 491	svs clk 491	0.165	0.000	0	430480	2 034	-0.253	-289 459	10007	430480	0.000	Yes	Partial False Path
sys_clk_245	sys_clk_491	0.916	0.000	0	3292	2.034	-0.383	-271.798	2022	3292	0.000	Yes	Partial False Patl
sys_clk_491	sys_clk_245	0.420	0.000	0	3751	2.034	-0.206	-254.735	2687	3751	0.000	Yes	Partial False Pati
svs_clk_245	sys_clk_122	0.461	0.000	0	1125	4.069	-0.224	-73.057	619	1125	0.000	Yes	Partial False Patl
sys_clk_245	sys_clk_61_44	10.444	0.000	0	3134	16.276	-0.234	-70.723	868	3134	0.000	Yes	Partial False Path
sys_clk_245	sys_clk_245	0.255	0.000	0	238146	4.069	-0.234	-46.686	1802	238146	0.000	Yes	Partial False Path
sys_clk_30_72	sys_clk_61_44	9.222	0.000	0	5348	16.276	-0.430	-41.222	5 4 5	5348	0.000	Yes	Partial False Path
svs_clk_30_72	svs_clk_122	27.186	0.000	0	12908	32 552	-0.356	-33,333	415	12908	0.000	Yes	Partial False Path
svs_clk_122	svs_clk_245	0.540	0.000	0	10781	4 069	-0.185	-21.341	580	10781	0.000	Yes	Timed
sys_clk_122	sys_clk_61_44	10.212	0.000	0	1083	16.276	-0.196	-14.464	172	1083	0.000	Yes	Timed

- 4 clocks identified with tightest setup requirement and highest THS violations
- Improve clock tree and apply CLOCK\_DELAY\_GROUP



### **Block RAM**

Guidelines for Paths from/to BRAM

Implement smaller memories in the LUTRAM (<8kb)</li>

- Avoid high fanout signals from/to big blocks
  - Replicate in RTL based on hierarchy with KEEP attribute
  - Advanced replication features coming in 2016.3 across the flow
- Use low-logic-level paths to drive control signals
  - Replicate address bus as needed or use PhysOpt
  - Recommend cascade\_height attribute for performance/power

```
(* ram_style = "distributed" *)
   [31:0] mem [(2**8)-1:0]:
   [7:0] addr_reg;
. . .
```

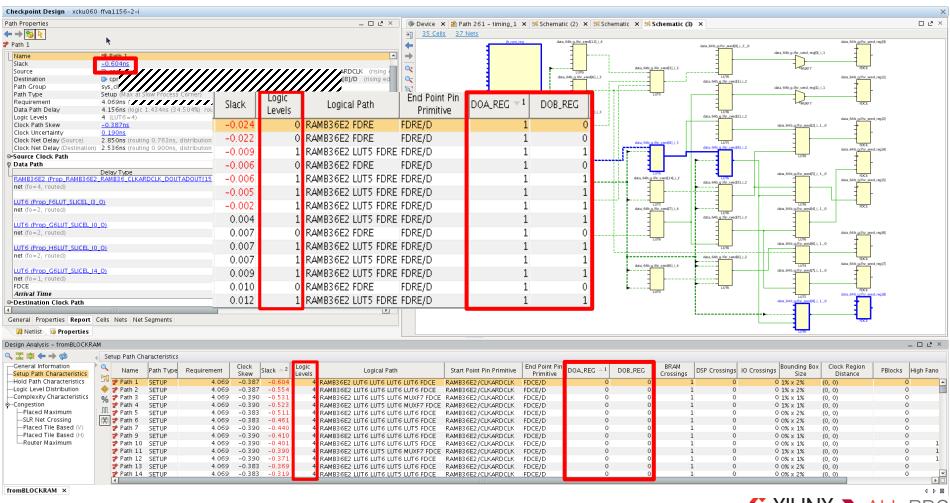
```
(* ram_style = "block".
cascade_height = 1/8/32 *
reg [31:0] mem [(2**15)-1:0];
   [14:0] addr_reg:
```

- Enable output register for high performance designs
  - Register should be either inside or directly connected to output pin
- Analyze block RAM paths with report\_design\_analysis

```
report design analysis -of timing paths [get timing paths \
-from/to [get cells -hier -filter {PRIMITIVE GROUP==BLOCKRAM}] \
-max 1000 -nworst 16 -unique pins] -file <from><to>BLOCKRAM.rpt
Page 22
```

# Real Customer Case Study - Unregistered RAMB output

Many paths from same RAMB had 4+ levels of logic @ 250MHz



# Complexity and Congestion

## **Design Complexity**

Analyze with report\_design\_analysis -complexity

- Complexity is measure of logic interconnection
- Complex modules may be down a few levels of hierarchy
  - Change the 'Hierarchical depth' in the GUI

High Rent, average fanout on larger instances

High LUT6%, MUXF\* utilization

Instance   Modules   Rent   Average   Total   Instances   LUT1   LUT2   LUT3   LUT4   LUT5   LUT5   Memory LUT   DSP   RAMB	MIIXE
witten   utop   0.51   2.73   38637   701   3507   2932   3016   3957   6128   1585   68   80   6128   61	
	1025
	^
№ 3 cpuEngine (or1200_top)         or1200_top         0.52         3.78         10308         137         598         879         792         928         2461         416         4         22	312
♦ IffEngine (fftTop) fftTop 124 3315 34 1151 56 42 12 116 769 64	0
Image: Property of the	331
% № 3 usbEngine1 (usbf_top_0) usbf_top_0 0.71 4.12 11414 241 845 913 1015 1374 1717 200 0 2:	331

- Goal
  - Identify difficult modules (>15k cells) with high rent (>0.65) and/or high average fanout (>4.0) before placement
- Solutions
  - Option 1: Try different synthesis options, adopt a bottom-up synthesis flow (OOC)
  - Option 2: Floorplan the difficult module

## Congestion

Analyze with report\_design\_analysis -congestion

## Physical regions with

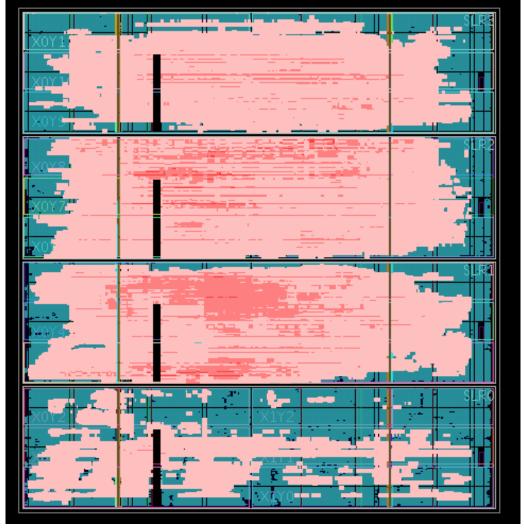
- High pin density
- High utilization of routing resources

## Placer is Congestion-aware:

- Balances congestion vs. wire length vs. timing
  - Cannot always eliminate congestion
  - > Cannot anticipate congestion due to hold fixing
  - > Timing estimation impacted by detours
- Reports congested areas seen by placer

## Router congestion

- Routing detours to handle congestion but impacts timing
- Reports areas with routing utilization close to 100%



"Smear" Maps

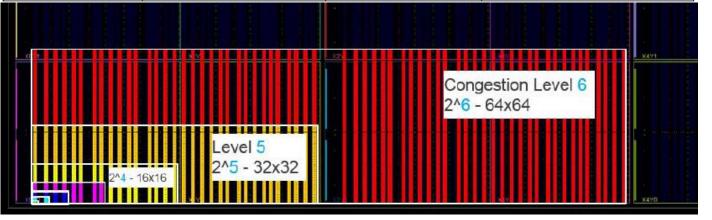
## Congestion Levels

### Definition

- Level corresponds to size of square area that contains congested routing
- Square size based on interconnect (INT\_XnYm) or CLB (CLE\_M\_XnYm)

### Congestion Level Ranges

Level	Area	Congestion	QoR Impact
1, 2	2x2, 4x4	None	None
3, 4	8x8, 16x16	Mild	Possible QOR degradation
5	32x32	Moderate	Likely QOR degradation
6	64x64	High	Difficulty routing
7,8	128x128, 256x256	Impossible	Likely unroutable



## Types of Congestion

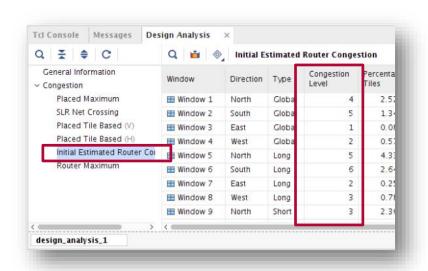
- Global: region of high utilization of routing resources
  - Demand for connections into/out of the region is high compared to supply
  - Common causes: High degree of LUT-combining, too many control sets, massive busses, poor floorplanning
  - UltraScale and UltraScale+ analysis often benefit from more detailed models: Short and Long
- Short: shortage of short-distance routes: SINGLEs, DOUBLEs, QUADs
  - More common in UltraScale
  - Common causes: dense concentration of MUXF or CARRY primitives
- Long: shortage of long-distance routes: HLONGs, VLONGs
  - More common in UltraScale+
  - Common causes: module with high Rent/average fanout, high big-block utilization with heavy connectivity, excessive SLR crossing nets

## Congestion Analysis: Recommended Approach

Check router warning messages

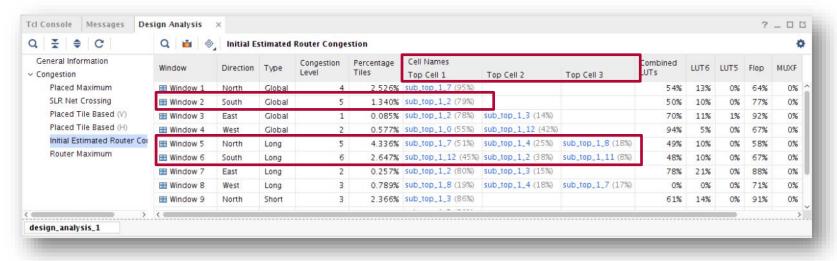
WARNING: [Route 35-447] Congestion is preventing the router from routing all nets.

- Run Report Design Analysis (IDE) or report\_design\_analysis -congestion
  - Focus on Initial Estimated Router Congestion
    - ➤ Most detailed report: Global, Long, and Short congestion
    - ➤ Congestion picture seen by the router after initial routing
    - ➤ Best opportunity to incorporate effective changes
  - Placed reports confirm congestion but are not as accurate
  - Router Maximum indicates whether congestion was resolved
- Check Initial Estimated Router congestion level
  - Level 4 and below usually not a problem
  - Level 5 some difficulty, can be resolved if isolated region
  - Level 6 major difficulty for routing, compile time and Fmax impacted
  - Level 7 and above usually impossible



## Congestion Analysis for UltraScale and UltraScale+

Identify hierarchical cells contributing to the congested region



- Check for congestion "symptoms" in those cells
  - report\_design\_analysis -congestion: check utilization columns
  - report\_design\_analysis -complexity: use -hierarchical\_depth to report cell complexity metrics
  - report\_qor\_suggestions: evaluate analysis and apply suggestions on problem areas
  - report\_utilization [-cells] : review utilization of problematic cells

## **Congestion Culprits**

- Congested regions may be due to netlist structure
  - Heavy MUXF usage can create shortage of connections
  - Long CARRY chains attract dense logic and connections
  - Large numbers of low-fanout control signals (reset, enable) consume routing resources
  - Excessive amount of complex logic cones (high Rent Exponent)
- Resynthesize the hierarchical cells contributing to congestion
  - Caution using strategies and options globally, possible wider effect on timing
  - Use block-level synthesis strategies to generate more suitable netlist
  - Out-of-context (OOC) flow can be used for solutions not yet handled by block-level synthesis

# Congestion Solutions (1/2)

- Disable LUT combining on a problem instance
  - Override existing LUT properties with empty value at beginning of Implementation

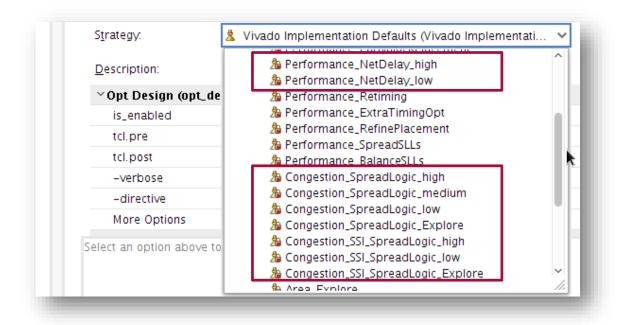
```
set_property HLUTNM "" get_cells -hier -filter {ref_name =~ LUT* && name =~ */inst/*}
```

- High utilization of MUXF and CARRY primitives may cause congestion
  - Use opt\_design to remap MUXF and CARRY primitives to LUTs
  - Global options -muxf\_remap and -carry\_remap: use with caution, may affect timing
  - Focus on cells in congested regions using cell properties MUXF\_REMAP and CARRY\_REMAP set property MUXF\_REMAP 1 [get\_cells inst\_a]; # opt\_design only remaps MUXF cells in inst\_a
- Further reduce logic levels to minimize timing impact
  - opt\_design -remap: affects entire design
  - Can use DONT\_TOUCH for surgical optimization

```
set orig_dont_touch_cells [get_cells -hier -filter DONT_TOUCH] ; # remember which cells have DONT_TOUCH
set DONT_TOUCH 1 <all cells except those to be remapped> ; # isolate the cells for optimization
opt_design -remap ; # only optimize the isolated cells
set DONT_TOUCH 0 [get_cells -hier] ; # removed DONT_TOUCH from all cells
set DONT_TOUCH 1 $orig_dont_touch_cells ; # restore original DONT_TOUCHed cells
```

# Congestion Solution (2/2)

Suggested Implementation Strategies



UltraScale+: NetDelay\* are good choices Less Short congestion, more Long congestion

UltraScale: SpreadLogic\* are good choices

See UG904 Appendix C for directives used in each strategy

# report\_qor\_suggestions (RQS)



- Analyzes design QoR issues and recommends solutions
  - Can be run at any design stage
  - Reports design issues against categories of different QoR checks
  - Suggests actions to take for each design issue
  - Provides turnkey XDC and Tcl files containing suggested fixes
- Next phase in progressing toward automated timing closure solutions
  - Current format: a text report with detailed descriptions and supplementary files
  - Future release: interactive GUI report
  - Eventual goal: Design Closure Wizard

Summary



iming\_summary
ming
locks (Note: Tcl only)
lock\_networks
lock\_interaction
dc
ethodology
esign\_analysis
ontrol\_sets
igh\_fanout\_nets

## Additional Resources

- UG 949
  UG 903
  All available now in the resource widget!
- **UG** 1231
- UltraFast Design Methodology for the Vivado Design Suite Introduction and Overview
- Constraint Explosion Quick Take Video
- UltraFast Vivado Design Methodology for Timing Closure
- <u>www.xilinx.com/vivado</u>

Q&A



