

SILEXICA 

SLX FPGA

Accelerate the journey from C/C++ to FPGA

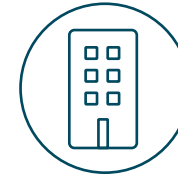
Silexica Facts



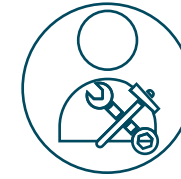
Est. 2014 after a decade of research



Team of world leading software and hardware experts



60 people worldwide, engineering HQ in Germany



3 offices and worldwide local support engineers



High-Level Synthesis Benefits

- High-level synthesis (HLS) provides C/C++ based FPGA design, benefits include
 - Faster implementation
 - Faster verification
 - Flexible design re-use
 - Faster design space exploration
- However, there are challenges...



High-Level Synthesis – First Impressions

- Can't compile – lots of synthesizability errors
- Slow performance and/or bloated area
- Difficult to detect parallelism and remove parallelism blockers
- Time consuming, iterative manual pragma optimization/insertion

Silexica SLX FPGA

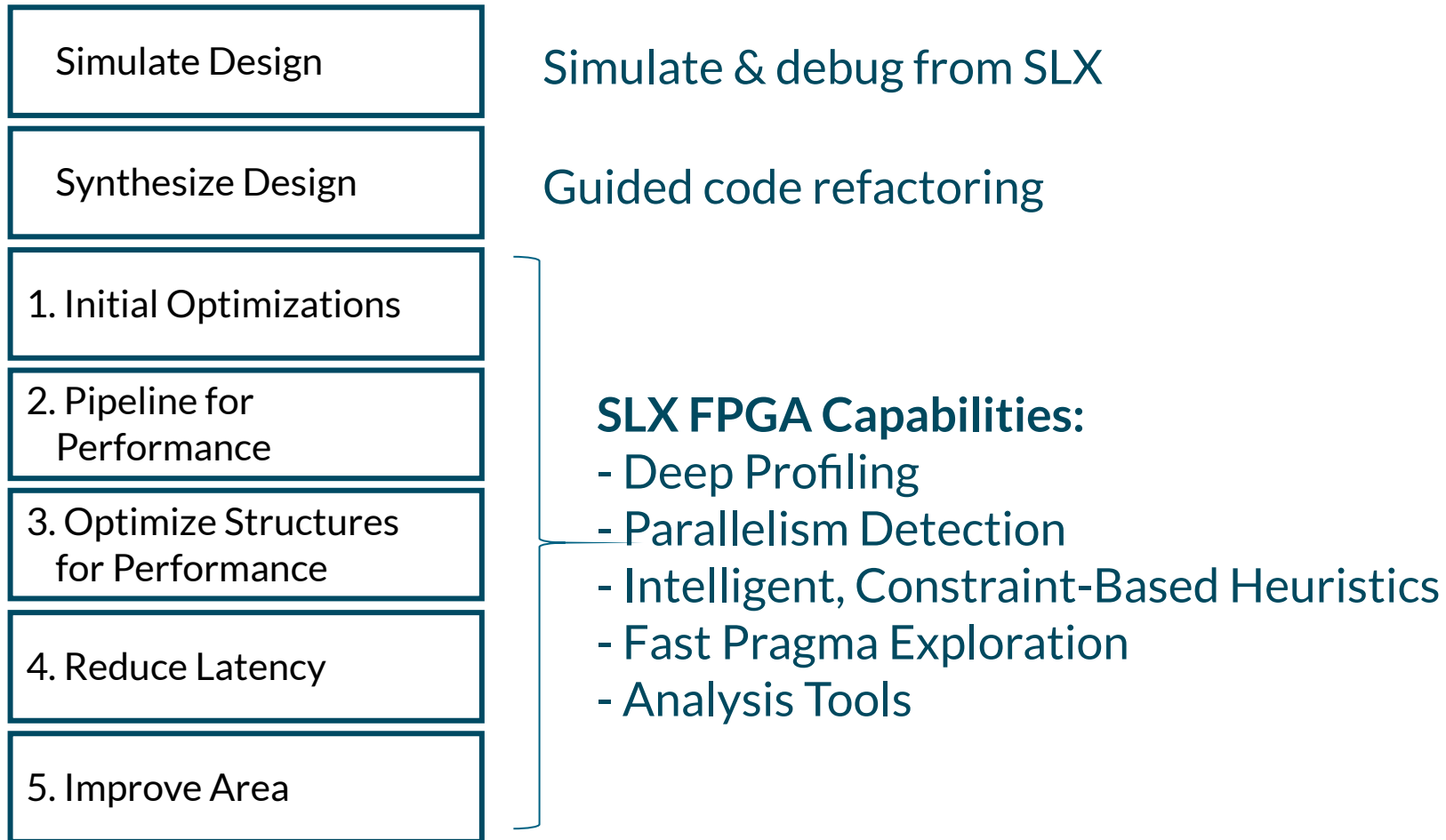
- SLX FPGA sits on top of HLS compiler
 - Prepares the C/C++ code for optimum HLS results
 - Takes the guesswork out of using HLS
- Removes the roadblocks in HLS adoption
 - Non-synthesizable C/C++ code
 - Finding parallelism
 - Poor performance and bloated area
- **HW engineers:** Get SW guidance needed
- **SW engineers:** Get parallelism/HW guidance



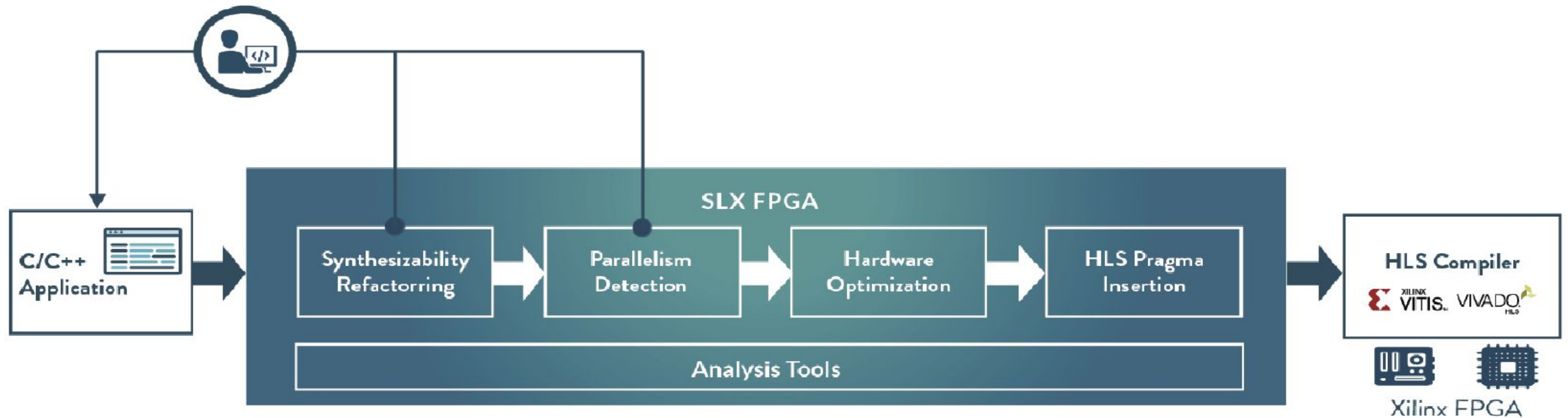
Vivado HLS Optimization Methodology Guide

Simulate Design	- Validate the C function
Synthesize Design	- Baseline design
1. Initial Optimizations	- Define interfaces (and data packing) - Define loop trip counts
2. Pipeline for Performance	- Pipeline and dataflow
3. Optimize Structures for Performance	- Partition memories and ports - Remove false dependencies
4. Reduce Latency	- Optionally specify latency requirements
5. Improve Area	- Optionally recover resources through sharing

SLX FPGA Assists at Every Stage



SLX FPGA - Automated Workflow for HLS



Synthesizability Refactoring

Code Refactoring for Synthesizability

- Not all C/C++ code is compatible with HLS
 - HLS has unique coding standards that must be followed
 - Not trivial – must become fluent in coding C/C++ for HLS
- SLX FPGA helps users refactor code for synthesizability
 - SLX FPGA identifies non-synthesizable functions code
 - Automatically refactors code for some common functions
 - Provides guided refactoring with hints on how to rewrite code



Automatic Code Refactoring

Automatic
refactoring
of well-known
functions

Direct link to
generated
code

The screenshot displays the Silexica IDE interface. The top editor pane shows the source code for `workshop_fpga.c`. A function `hwscale_accum` is defined, which takes `fp_complex` arrays and a scaling factor. The function contains a loop that is annotated with HLS directives. The bottom pane shows the generated code, which includes a filter and a list of transformations applied to the function.

```
void hwscale_accum(fp_complex src[SAMPLES], fp_complex dst[SAMPLES],  
                  fp_complex *scaling, unsigned int size) {  
    unsigned int shift = slx_fpga_rand() % 3 + 2;  
    for(int i = 0; i < SAMPLES; i++) {  
        // Loop annotated with HLS directives by SLX  
        #pragma HLS loop_tripcount min=1024 max=1024  
        // Loop annotated with HLS directives by SLX  
        #pragma HLS unroll factor=128 skip_exit_check
```

Filter

- for the function (swscale_accum)
- inline pragma inserted for function compMulScale
- Wrapper - Non-synthesizable call to function: rand has been replaced with call to slx_fpga_rand
- HLS loop_tripcount pragma reporting 1024 iterations, inserted for the loop
- HLS unroll pragma with unroll factor of 128 and skip_exit_check inserted for the loop

Guided Code Refactoring

Highlighting of
non-synthesizable
code (access to
pointer)

The screenshot shows the Silexica IDE interface. The main editor window displays the code for `workshop_fpga.c`. The line `fp_complex *scaled_vector_sw = malloc(SAMPLES * sizeof(fp_complex));` is highlighted in green, indicating it is non-synthesizable. Below the editor, the 'SLX' panel shows a list of errors. The error '44:44] unsupported c/c++ library function 'malloc'' is highlighted in orange. The right panel, titled 'Help', provides guidance for code re-writing. It shows two examples: 'Non-synthesizable memory allocation' using `calloc`, `free`, and `malloc`, and 'Synthesizable memory allocation' using static arrays.

```
int main() {
    for(int i = 0; i < ITER; i++) {
        fp_complex *scaled_vector_sw = malloc(SAMPLES * sizeof(fp_complex));
        fp_complex scaled_vector_hw[SAMPLES];
#ifdef SLX_DEBUG
        printf("*** Original: ");
        dump(data_vector, SAMPLES);
#endif
        swscale_accum(data_vector, scaled_vector_sw, &data_vector);
        hwscale_accum(data_vector, scaled_vector_hw, &data_vector);
    }
}
```

Non-synthesizable memory allocation:

```
int *a = calloc(ARRLEN, sizeof(int));
int *b;
b = (int *) malloc(ARRLEN * sizeof(int));
free(a);
free(b);
```

Synthesizable memory allocation:

```
int a[ARRLEN];
int *a = &a[0];
int b[ARRLEN];
int *b = &b[0];
```

Summarized
report to help
refactoring

Guidance for
code re-writing

Parallelism Detection

Parallelism Detection

- Identifying parallelism is difficult
 - Even more difficult if HLS user did not write algorithm
- SLX FPGA analyzes the applications and identifies parallelism patterns to implement in hardware
 - Identifies Data Level and Pipeline Level Parallelism
 - Also provides insights into parallelism blockers



Parallelism Detection

Source
highlighting for
parallel code

Configuration Editor - Function Mapping Edit workshop_fpga.c workshop_fpga.c

```
void hwscale_accum(fp_complex src[SAMPLES], fp_complex dst[SAMPLES],
                  fp_complex *scaling, unsigned int size) {
    unsigned int shift = rand() % 3 + 2;
    for(int i = 0; i < SAMPLES; i++) {
        fp_complex res = {0,0};
        res = compMulScale(&src[i], scaling, shift);
        dst[i] = res;
    }
}
```

Problems Console Properties SLX Hints Code Analysis Progress

Filter

Name	Status	Description	Local Sp	Global S	Exec. Pe	He
PARTITIONING		workshop_fpga.c [35:39] The loop has the following indu				
DLP	✓		1023.9	2.0		
HLS		../output/codegen/hls/spec/workshop_fpga.c [40:42] HL				
PARTITIONING		workshop_fpga.c [35:39] The loop will use 1024 workers.				
PARTITIONING		workshop_fpga.c [35:39] DLP parallelism finally selected				

Summarized
report to help
navigation

Parallelism Detection

Blocked
DLP

PLP

DLP

	Name	Status	Location	Description	Help
14	Loop		..\src\rng.cpp [83:95]		
15	HLS		..\output\codegen\hls\src\...	HLS loop_tripcount pragma reporting 312 iterations, inserted for the loop	?
16	DLP	✗			
17	PARTITIONING		..\src\rng.cpp [83:95]	The loop does not provide DLP because of loop-carried dependencies.	?
18	PARTITIONI...		..\src\blackScholes.cpp	Loop-carried dependency on variable mt_rng [WAW]	?
19	PARTITIONI...		..\src\rng.cpp [67:67]	Loop-carried dependency on variable tmp [RAW]	?
20	PARTITIONING		..\src\rng.cpp [83:95]	The loop (83:95) presents no beneficial DLP for FPGA	?
21	PLP	✓			
22	HLS		..\output\codegen\hls\src\...	HLS pipeline pragma inserted for the loop	?
23	PARTITIONING		..\src\rng.cpp [83:95]	PLP parallelism available for loop (83:95).	?
24	Loop		..\src\rng.cpp [85:94]		
25	HLS		..\output\codegen\hls\src\...	HLS loop_tripcount pragma reporting 2 iterations, inserted for the loop	?
26	PARTITIONING		..\src\rng.cpp [85:94]	The loop has the following induction variable:	?
27	PARTITIONI...		..\src\rng.cpp [85:85]	Induction variable: k	?
28	DLP	✓			
29	HLS		..\output\codegen\hls\src\...	HLS unroll pragma with unroll factor of 2 and skip exit check inserted for the loop	?
30	PARTITIONI...		..\src\rng.cpp [85:94]	DLP parallelism available for loop (85:94). Unroll factor can be 2.	?

Blockers
identified

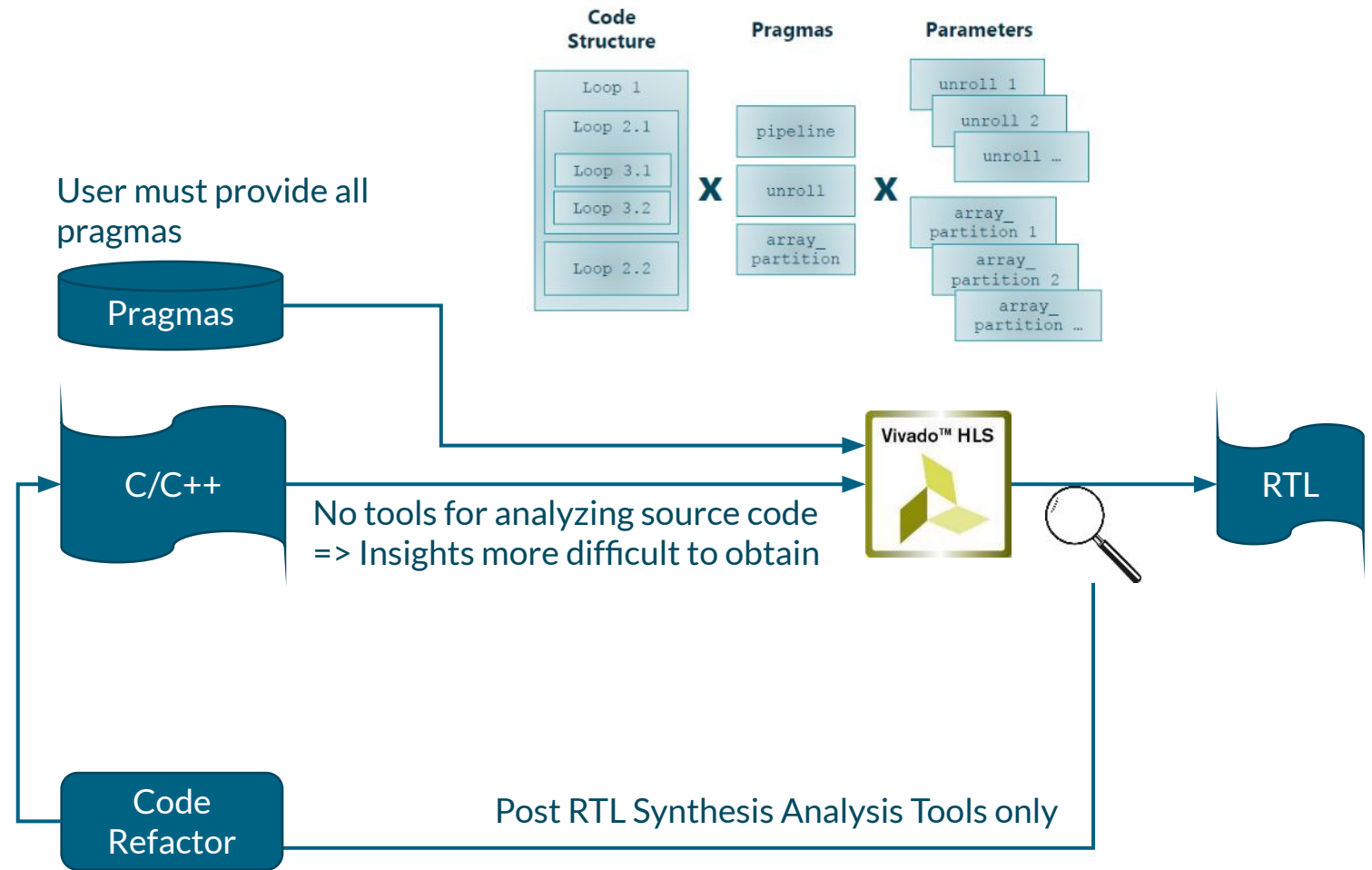
Automatic
tripcount
insertion

HW Optimization

Standard Vivado HLS Flow

Pragma Exploration is Tricky

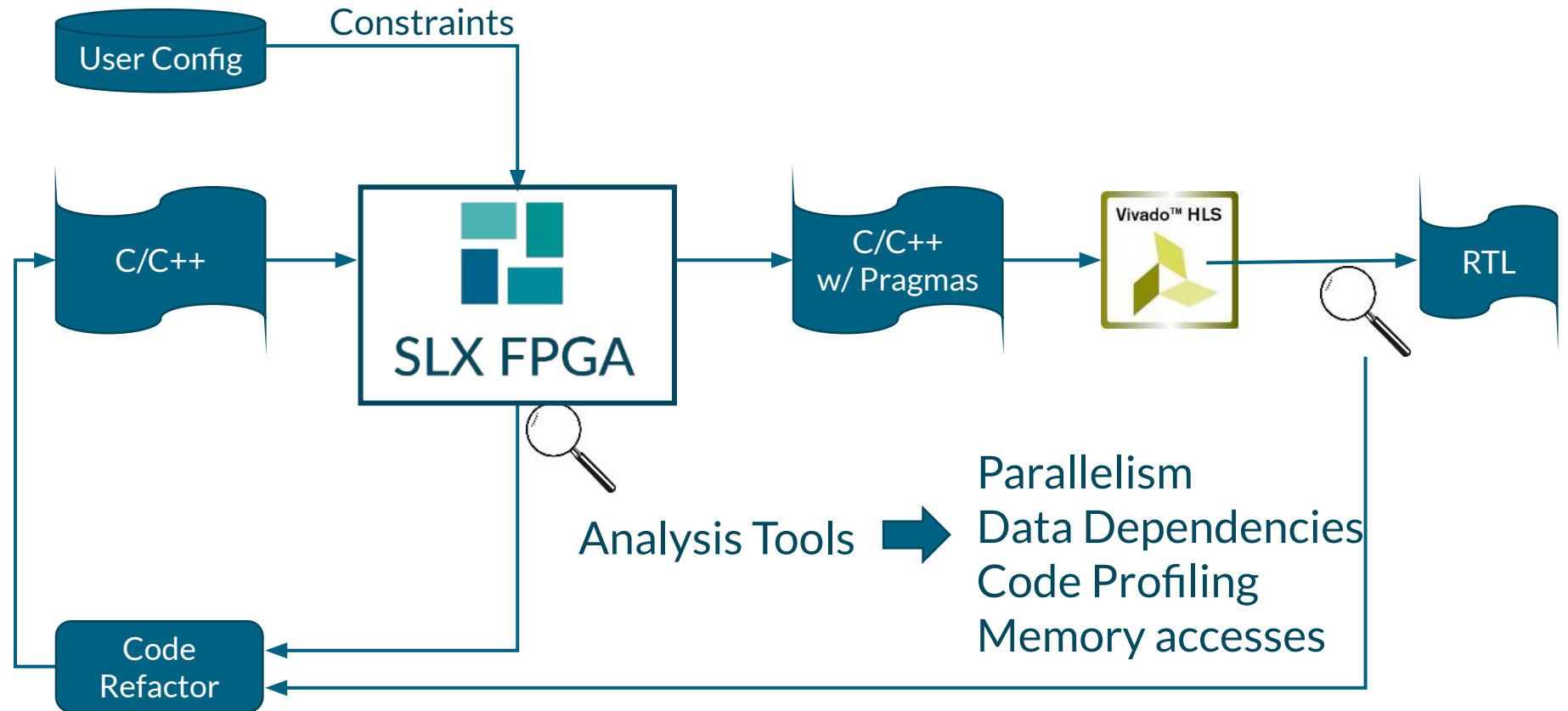
- Powerful, but requires detailed knowledge of code and Vivado HLS
- Even small pragma/parameter set leads to large design space
- Each combination needs to be synthesized with HLS
- Some combinations can lead to bloated implementations, extended synth times



Vivado/Vitis HLS + SLX FPGA Flow

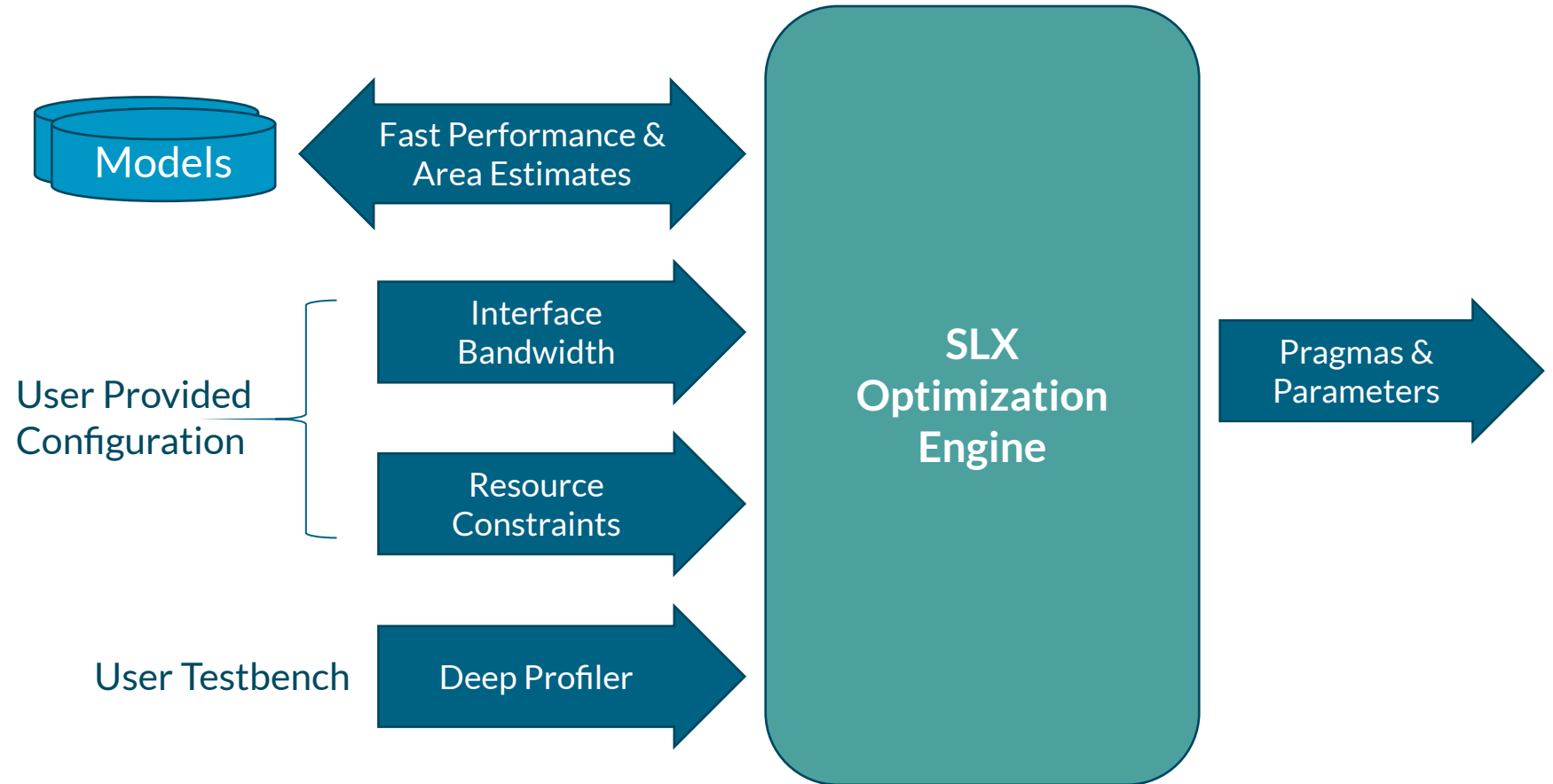
SLX FPGA

- Analyzes and optimizes design based on constraints
- Automatically inserts optimal pragmas into source code
- Analysis tools assist with code refactoring



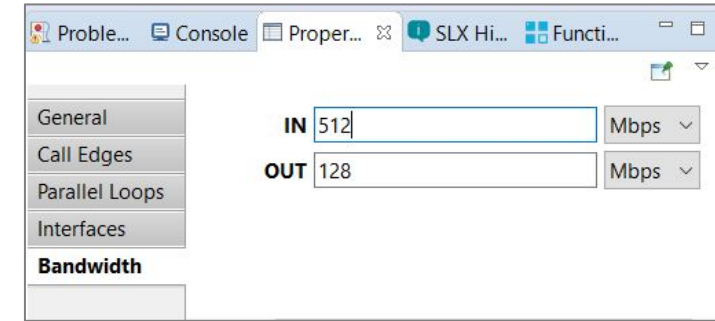
The SLX Optimization Engine

- Fully automated heuristics guided by deep profiling information
- Internal models enable fast architectural exploration
- Ability to explore impacts of different configuration

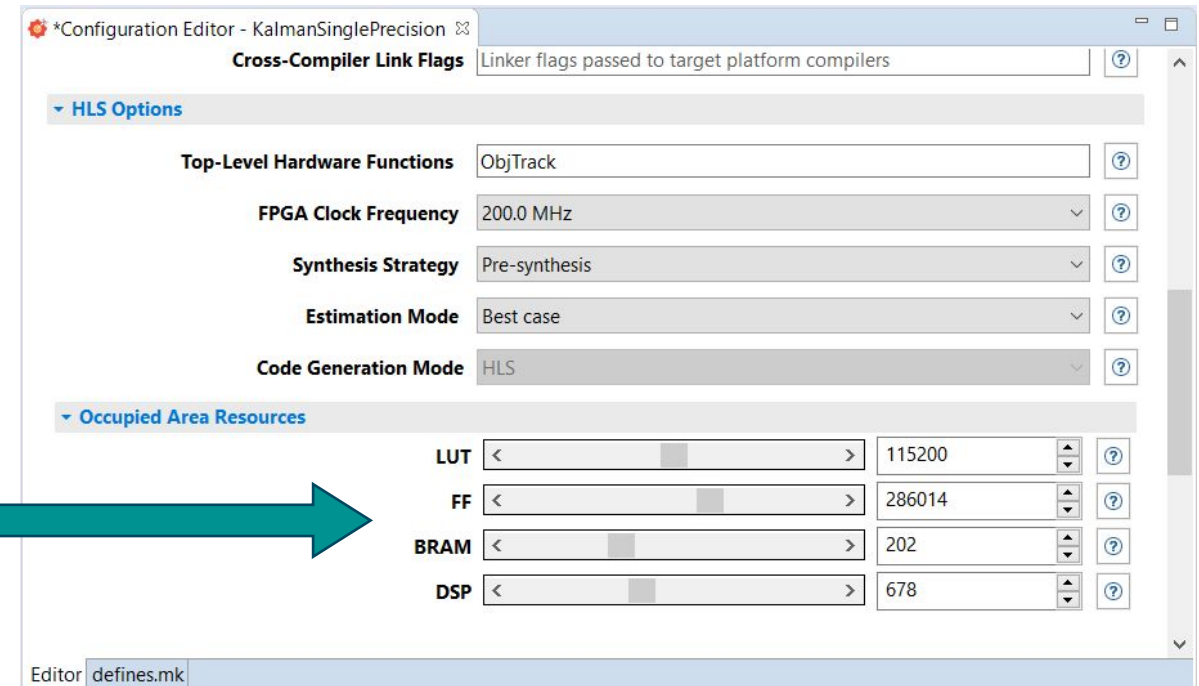


Constraints

Interface bandwidth

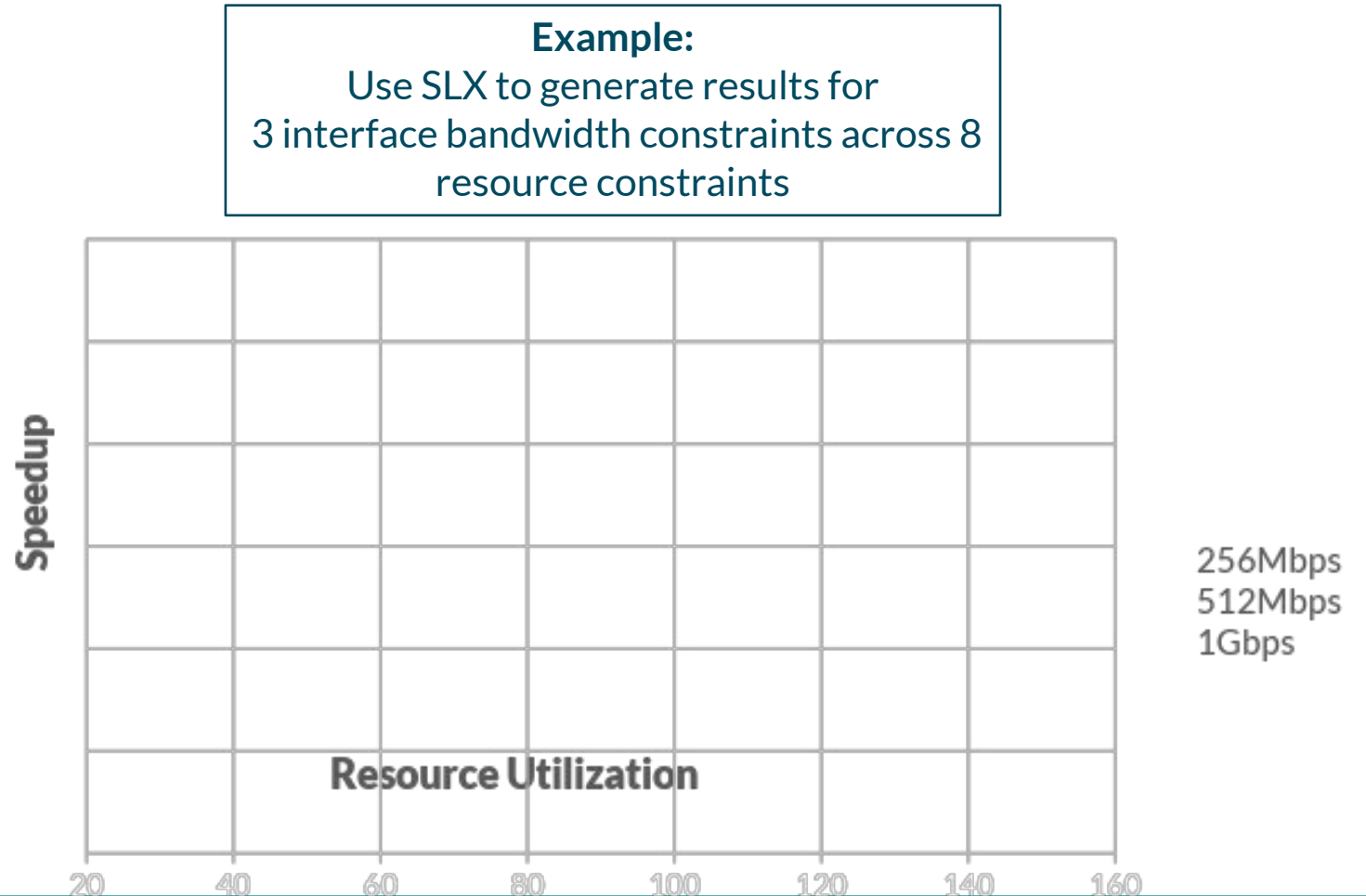


Resource constraints



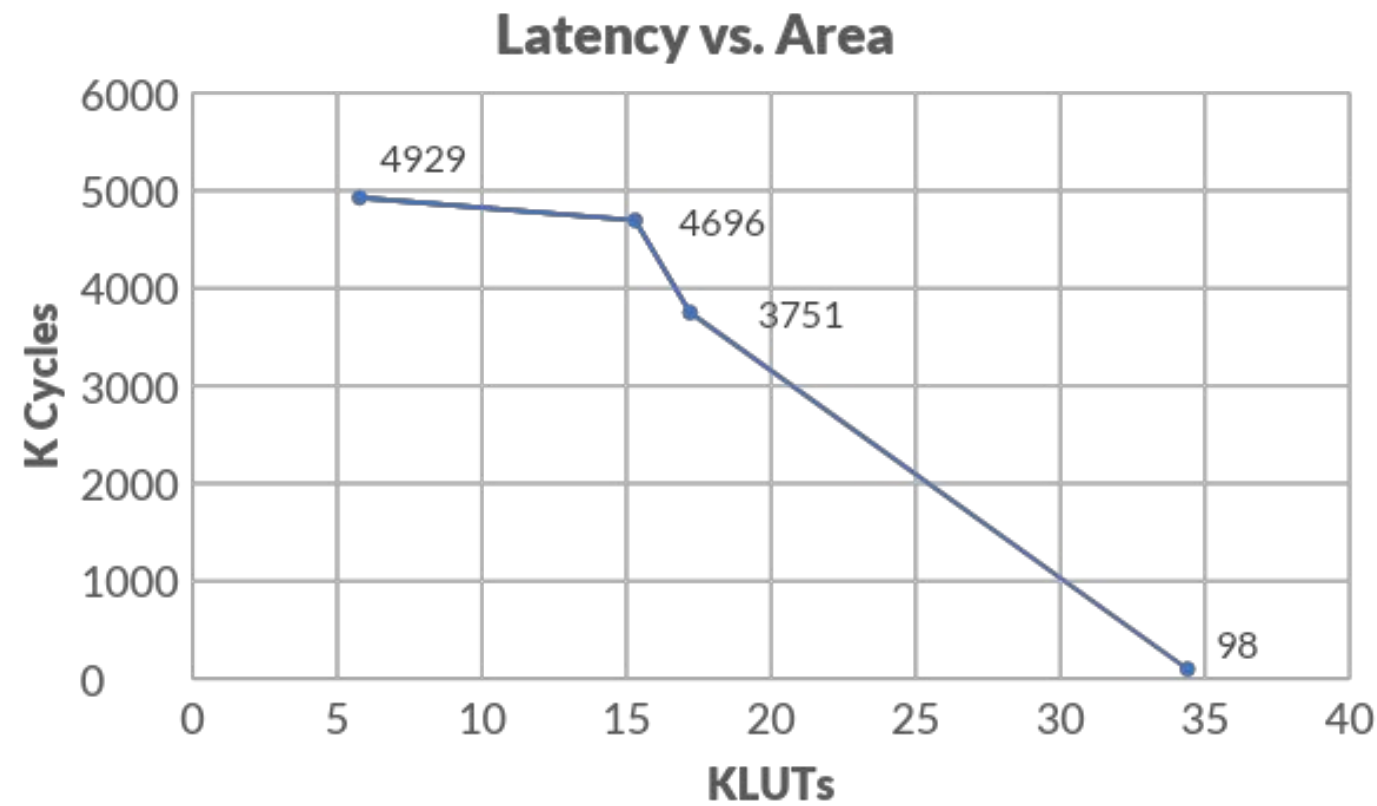
Design Space Exploration (DSE)

- Evaluate multiple user configuration parameters
- Enables area/performance trade-off analysis

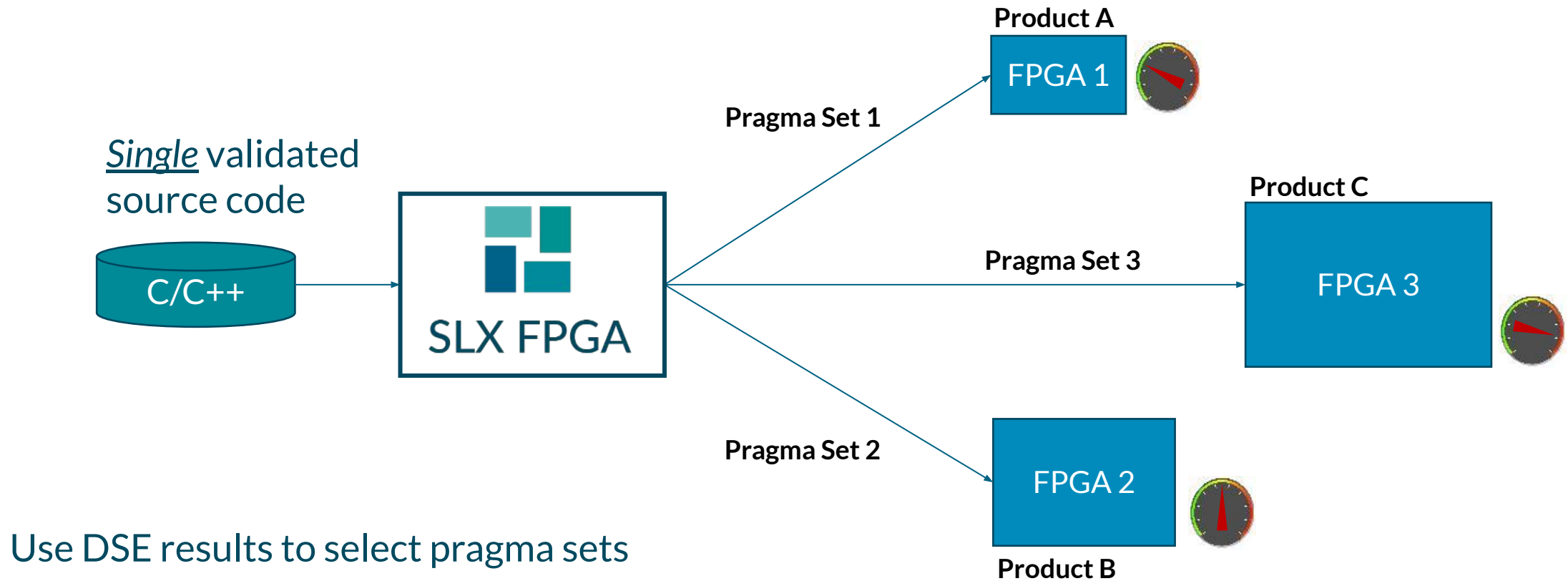


Kalman Filter Example

SLX Constraint (KLUTs)	Actual LUT usage from HLS (KLUTs)	Latency (1000s of clock cycles)
Tripcount only	5.8	4929
12	15.3	4696
24	17.2	3751
48	34.4	98
Unlimited	34.4	98



IP Re-use



Pragma Insertion

Automatic Pragma Insertion

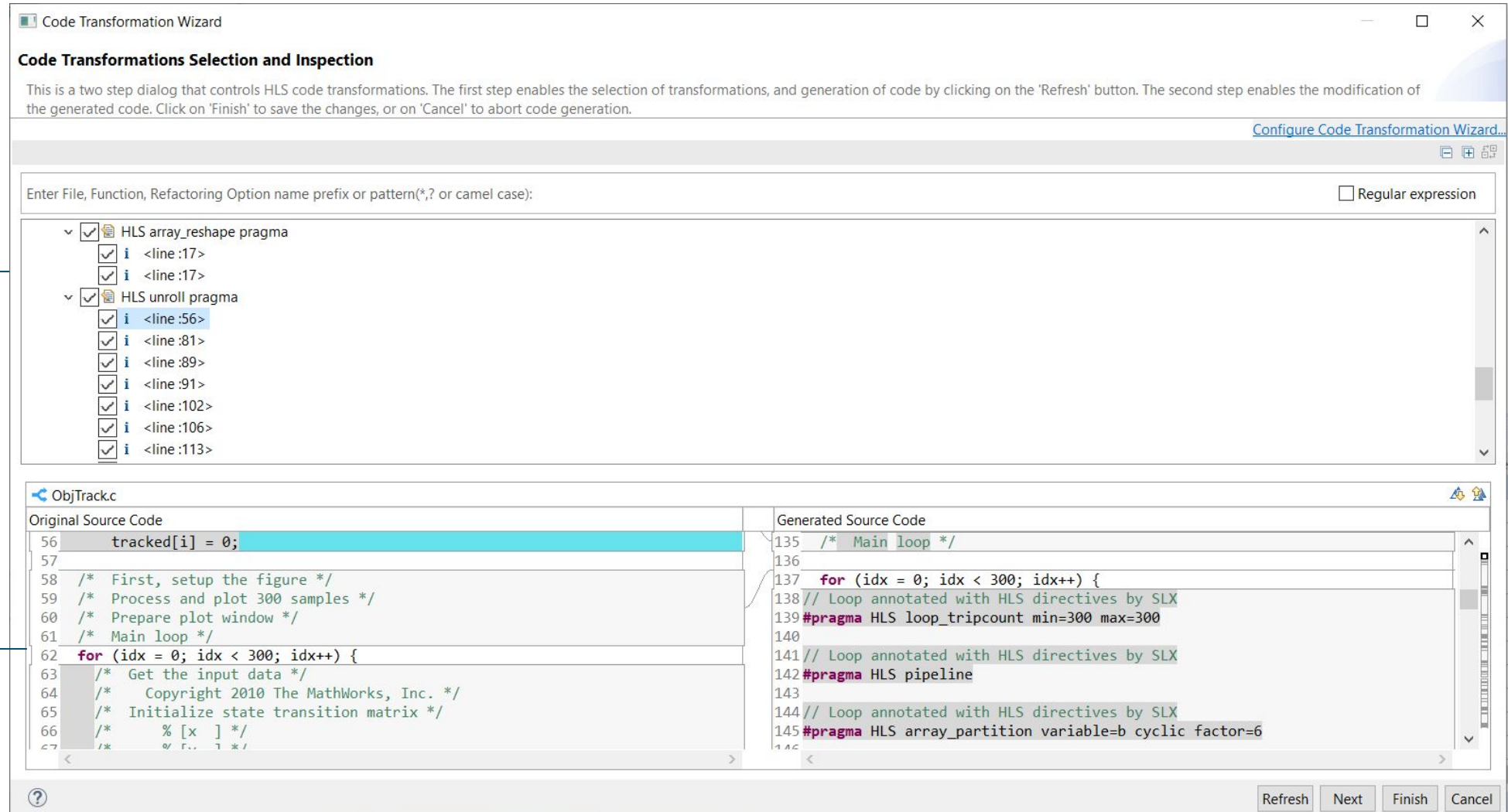
- HW optimization stage
 - Creates optimal pragma set
- Pragma Insertion Wizard
 - Inserts all generated pragmas into original source code
 - Allows designer full control over pragma insertion



SLX Code Transformation Wizard

Enable/disable individual or groups of pragmas

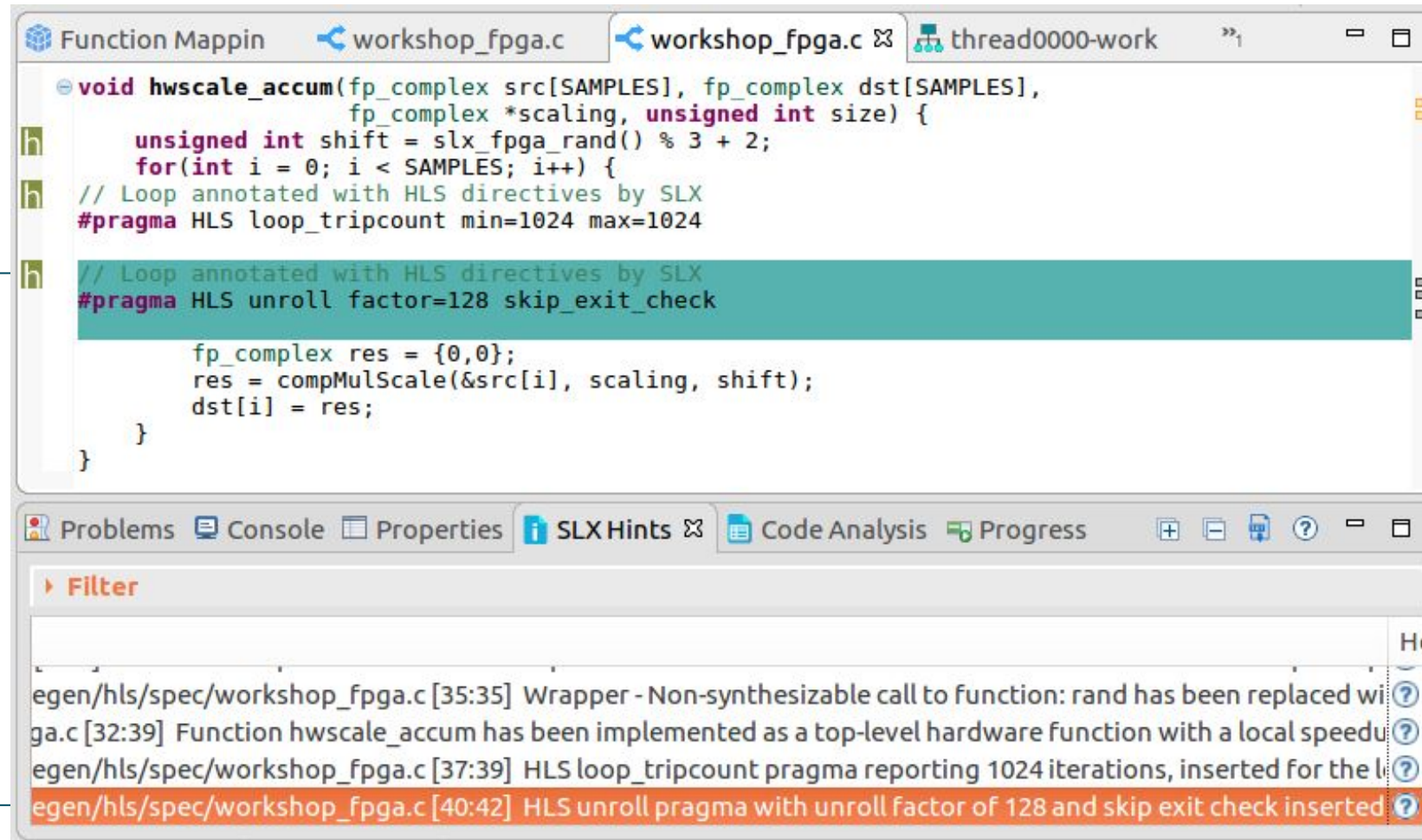
Original/Updated code side-by-side view



Automatic HLS Pragma Insertion

Automatically
annotate the code
with pragmas

Direct link to
generated file



The screenshot shows an IDE window with the following tabs: Function Mapping, workshop_fpga.c, workshop_fpga.c, and thread0000-work. The main editor displays the following C code:

```
void hwscale_accum(fp_complex src[SAMPLES], fp_complex dst[SAMPLES],  
                  fp_complex *scaling, unsigned int size) {  
    unsigned int shift = slx_fpga_rand() % 3 + 2;  
    for(int i = 0; i < SAMPLES; i++) {  
        // Loop annotated with HLS directives by SLX  
        #pragma HLS loop_tripcount min=1024 max=1024  
  
        // Loop annotated with HLS directives by SLX  
        #pragma HLS unroll factor=128 skip_exit_check  
  
        fp_complex res = {0,0};  
        res = compMulScale(&src[i], scaling, shift);  
        dst[i] = res;  
    }  
}
```

Below the code editor, there is a toolbar with icons for Problems, Console, Properties, SLX Hints, Code Analysis, and Progress. Below the toolbar is a list of generated files with a search filter. The list contains the following entries:

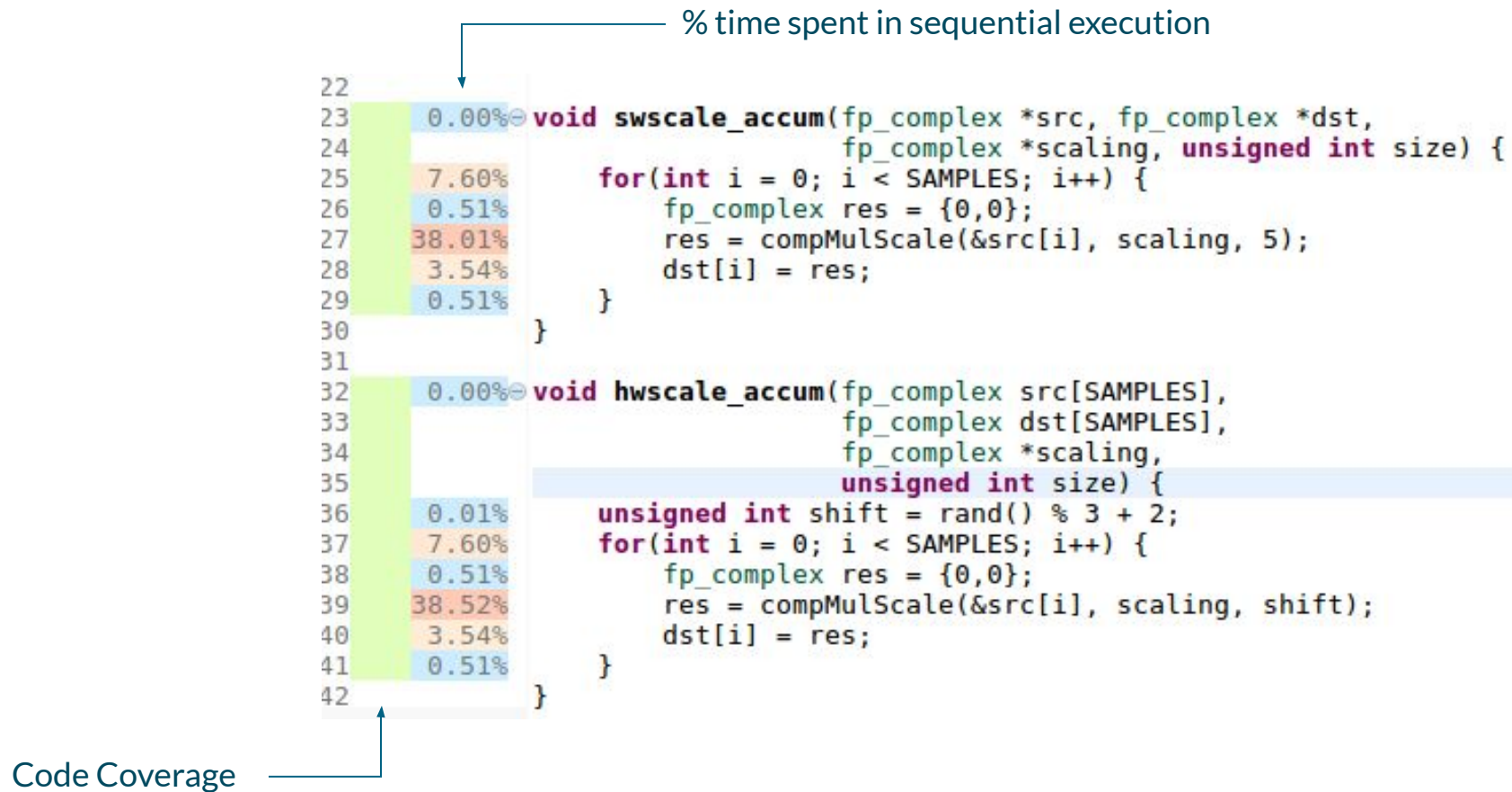
- egen/hls/spec/workshop_fpga.c [35:35] Wrapper - Non-synthesizable call to function: rand has been replaced with slx_fpga_rand
- egen/hls/spec/workshop_fpga.c [32:39] Function hwscale_accum has been implemented as a top-level hardware function with a local speedup
- egen/hls/spec/workshop_fpga.c [37:39] HLS loop_tripcount pragma reporting 1024 iterations, inserted for the loop
- egen/hls/spec/workshop_fpga.c [40:42] HLS unroll pragma with unroll factor of 128 and skip exit check inserted

Deep Profiling and Analysis Tools

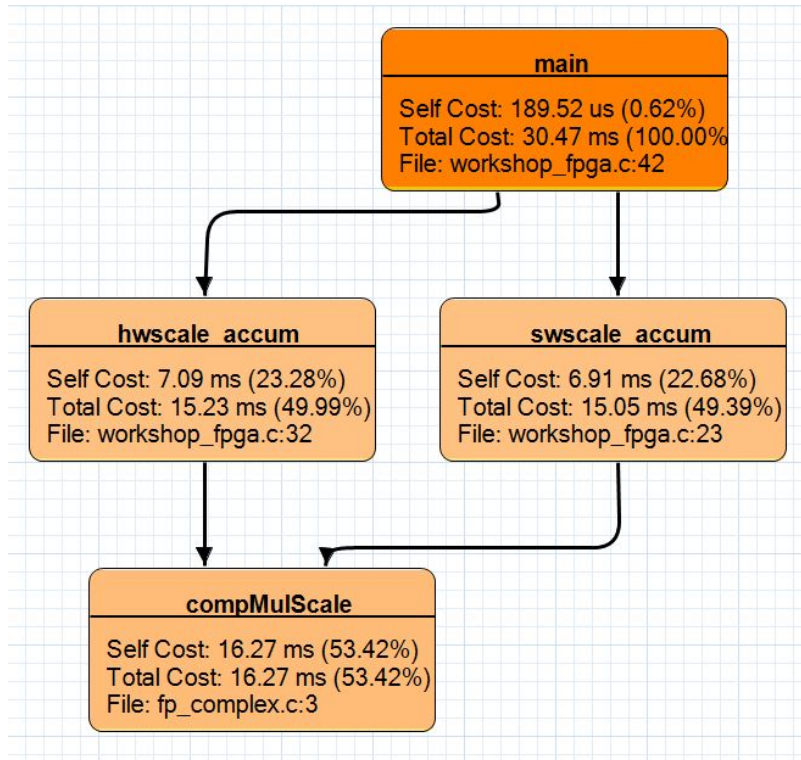
Hardware Aware Coding

- Pragmas can only go so far
 - Modifying the original code to be more hardware aware can lead to better performance and area
- Hardware aware refactoring requires insights into the algorithm.
- SLX Analysis tools provide detailed insights on the source code:
 - Code and Function Profiling
 - DLP and PLP detection
 - Data dependency detection
 - Software Call Graphs
 - Hotspot detection
 - Memory/variable analysis
 - Code analysis graphs

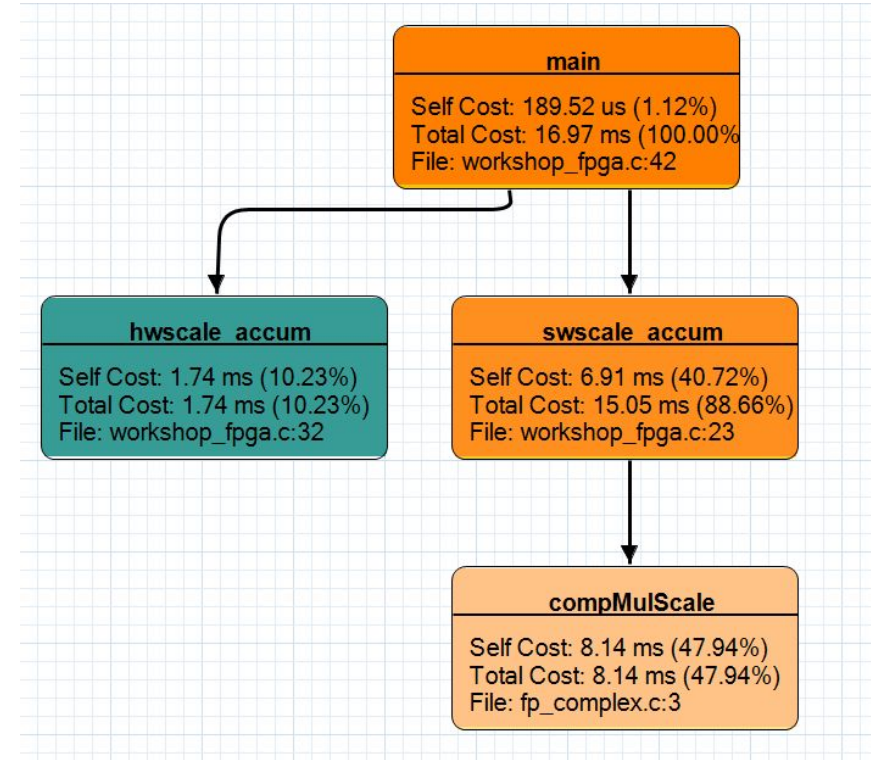
Code Profiling



Software Call Graphs with Profiling

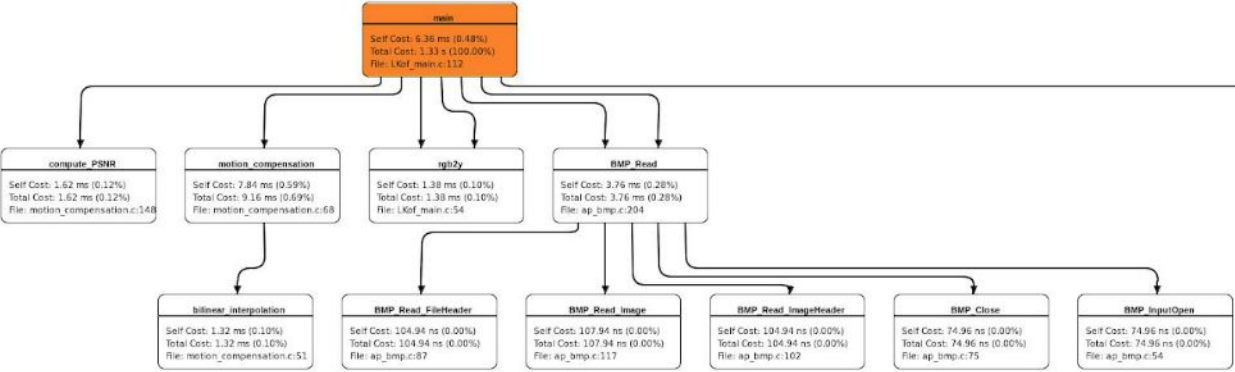


Pre-Optimization

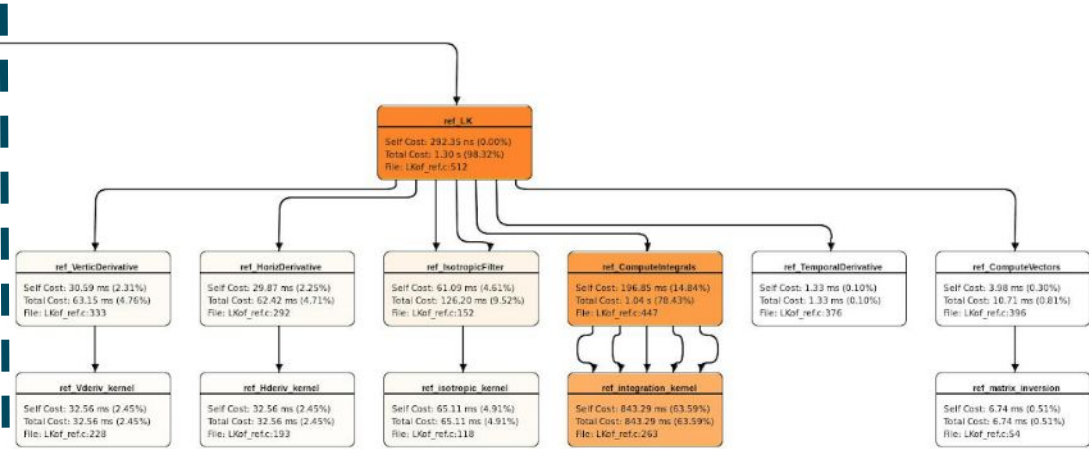


Post-Optimization

Hotspot Detection



Testbench Code



Synthesizable Code

Memory Analysis

Integrated source code highlighting

```
1 #include "fp_complex.h"
2
3 fp_complex compMulScale(fp_complex *a, fp_complex *b, int scale)
4 {
5     fp_complex c;
6     // c.real = fp_sub(fp_mul_scale(a.real, b.real, scale), fp_mul_scale(a.imag, b.imag, scale));
7     fp2 temp = ((fp2) (a->real)) * ((fp2) (b->real));
8     fp temp1 = (int)((temp >> (N + scale)));
9     temp = ((fp2) (a->imag)) * ((fp2) (b->imag));
10    fp temp2 = (int)((temp >> (N + scale)));
11    c.real = temp1 - temp2;
12    // c.imag = fp_add(fp_mul_scale(a.real, b.imag, scale), fp_mul_scale(a.imag, b.real, scale));
13    temp = ((fp2) (a->real)) * ((fp2) (b->imag));
```

Problems Console Properties SLX Hints Code analysis Memory Analysis

Variables

	Name	Type	File	Function	Line	Size (Bytes)	Reads	Writes	Total Access	Access
1	Global									
2	data_vector	fp_complex[1024]	data.c		5	8192	819200 (10.13%)	0 (0.00%)	819200 (6.78%)	
3	scaling	fp_complex	data.c		3	8	819200 (10.13%)	0 (0.00%)	819200 (6.78%)	
4	Heap									
5	Local									
6	temp	fp2	fp_complex.c	compMulScale	7	8	819200 (10.13%)	819200 (20.51%)	1638400 (13.56%)	
7	temp		fp_complex.c	compMulScale	7	8	0 (0.00%)	204800 (5.13%)	204800 (1.70%)	
8	temp		fp_complex.c	compMulScale	8	8	204800 (2.54%)	0 (0.00%)	204800 (1.70%)	
9	temp		fp_complex.c	compMulScale	9	8	0 (0.00%)	204800 (5.13%)	204800 (1.70%)	
10	temp		fp_complex.c	compMulScale	10	8	204800 (2.54%)	0 (0.00%)	204800 (1.70%)	
11	temp		fp_complex.c	compMulScale	13	8	0 (0.00%)	204800 (5.13%)	204800 (1.70%)	
12	temp		fp_complex.c	compMulScale	14	8	204800 (2.54%)	0 (0.00%)	204800 (1.70%)	
13	temp		fp_complex.c	compMulScale	15	8	0 (0.00%)	204800 (5.13%)	204800 (1.70%)	
14	temp		fp_complex.c	compMulScale	16	8	204800 (2.54%)	0 (0.00%)	204800 (1.70%)	
15	a	fp_complex*	fp_complex.c	compMulScale	3	8	819200 (10.13%)	204800 (5.13%)	1024000 (8.48%)	
16	b	fp_complex*	fp_complex.c	compMulScale	3	8	819200 (10.13%)	204800 (5.13%)	1024000 (8.48%)	
17	scale	int	fp_complex.c	compMulScale	3	4	819200 (10.13%)	204800 (5.13%)	1024000 (8.48%)	

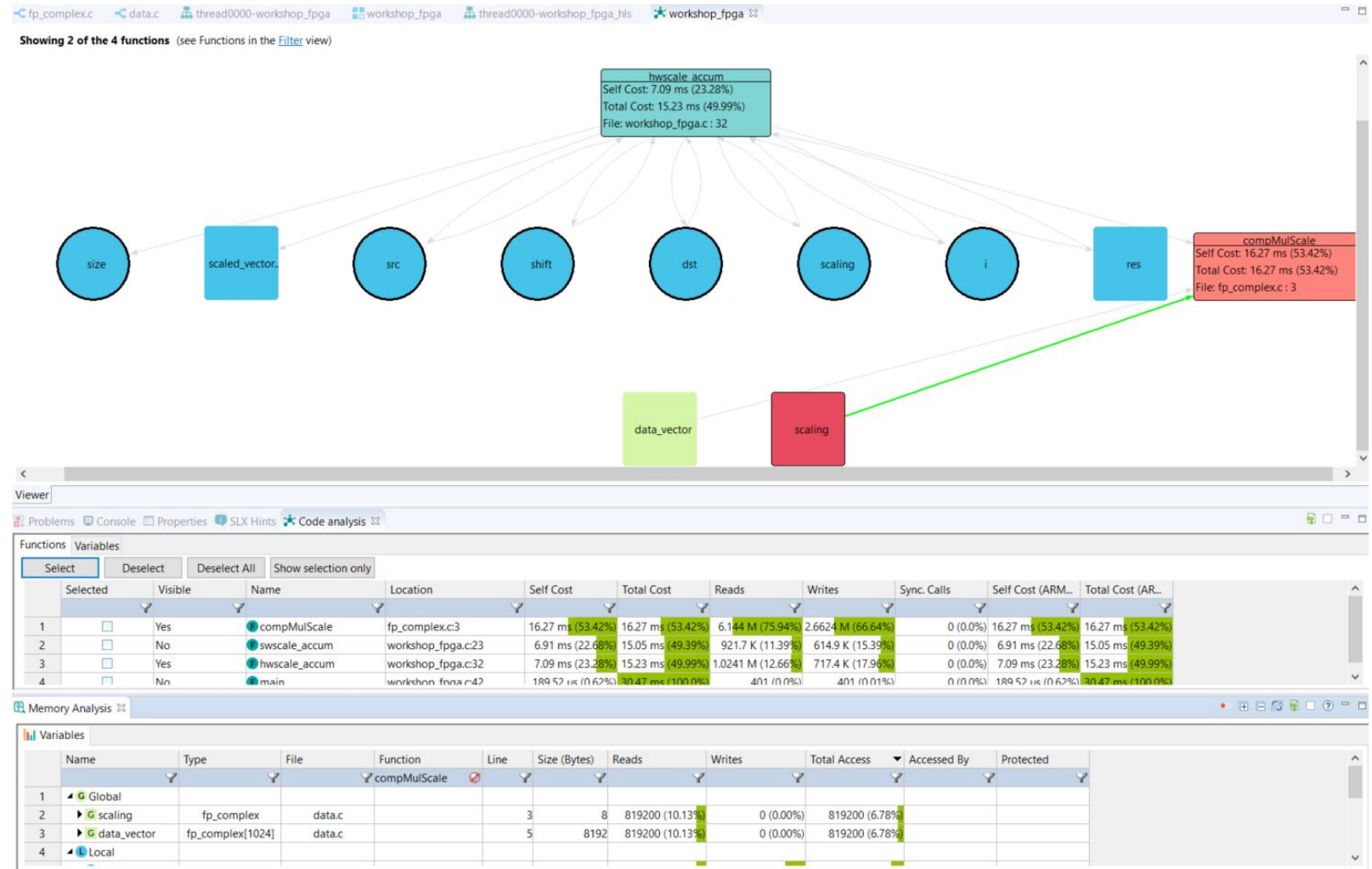
Access statistics

Location

Size

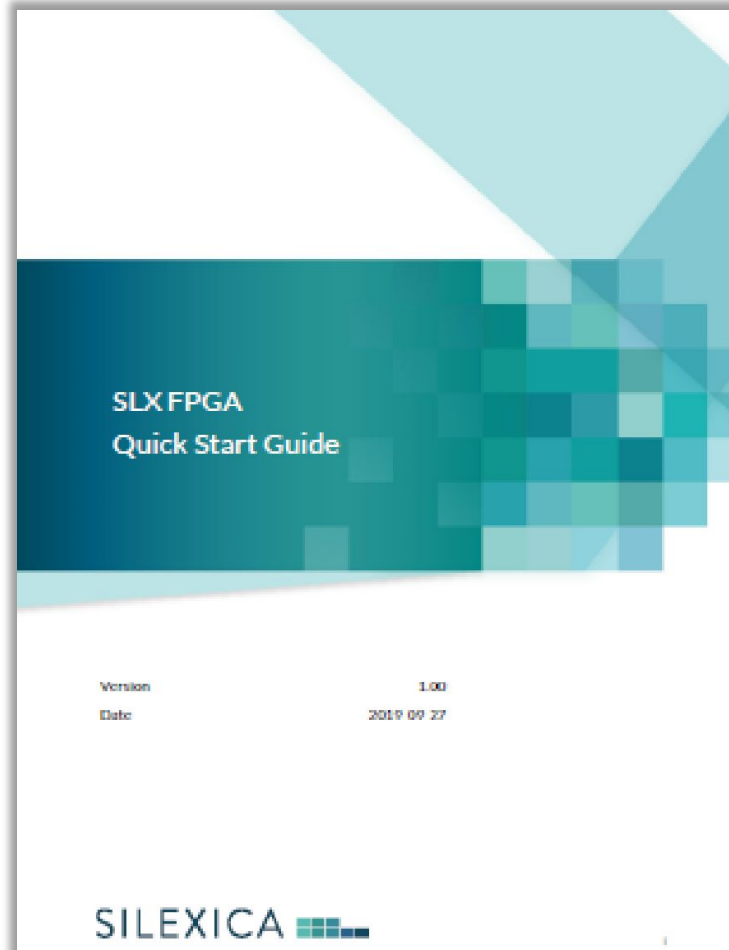
Code Analysis Graph

- Visual representation of relationship between variables and the functions that access them
- Filters allow control over the details in the graph
- Simultaneous variable and function access information

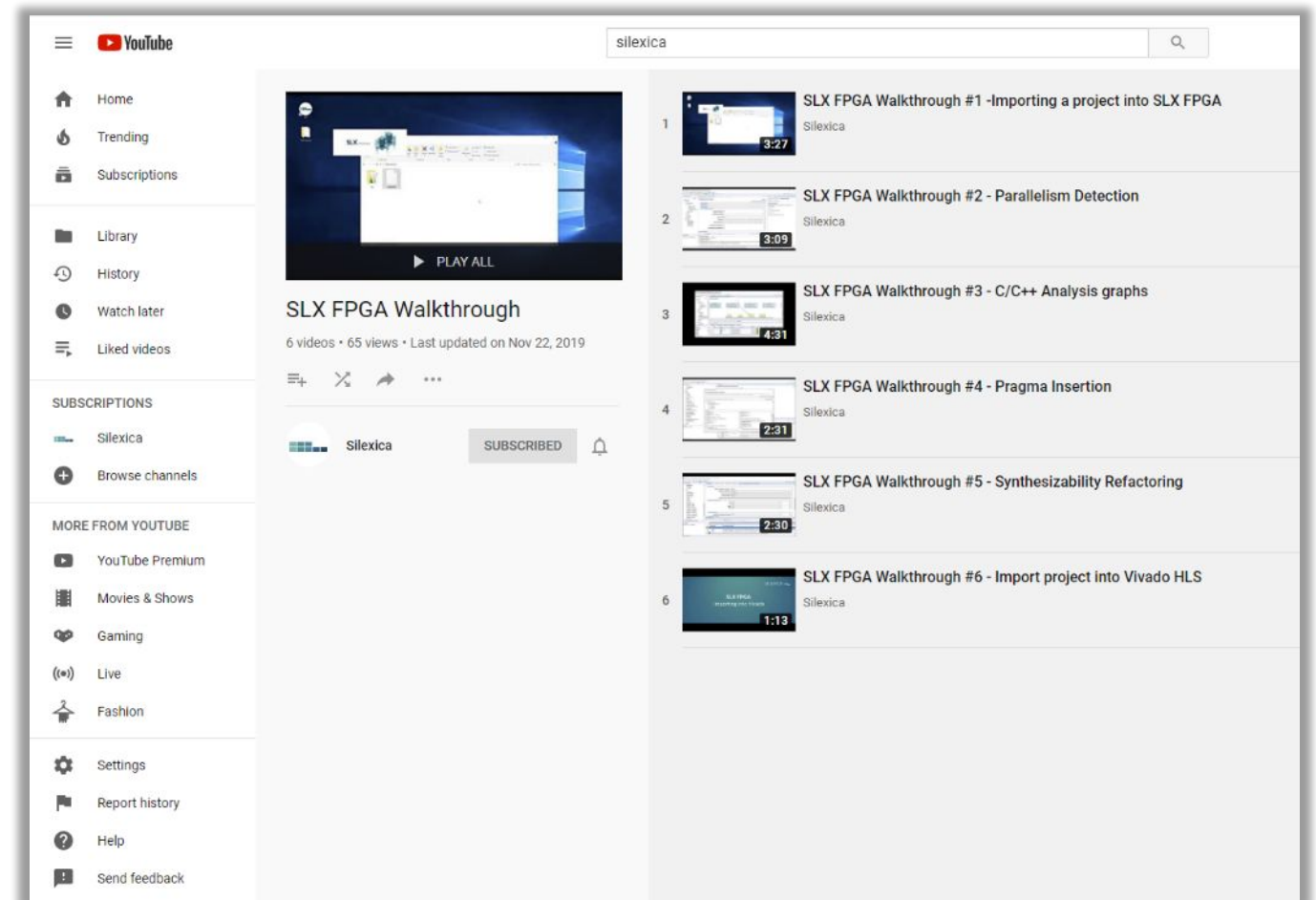


Collateral & Real World Results

SLX FPGA Getting Started

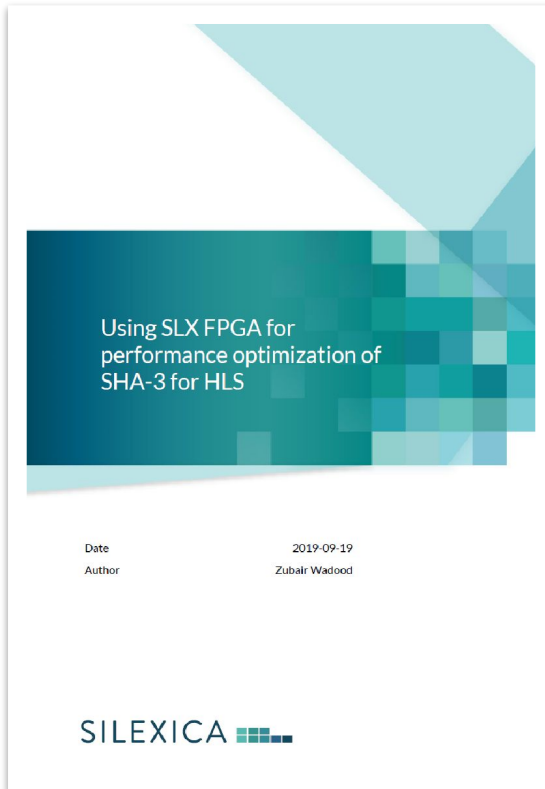


Quick Start Guide



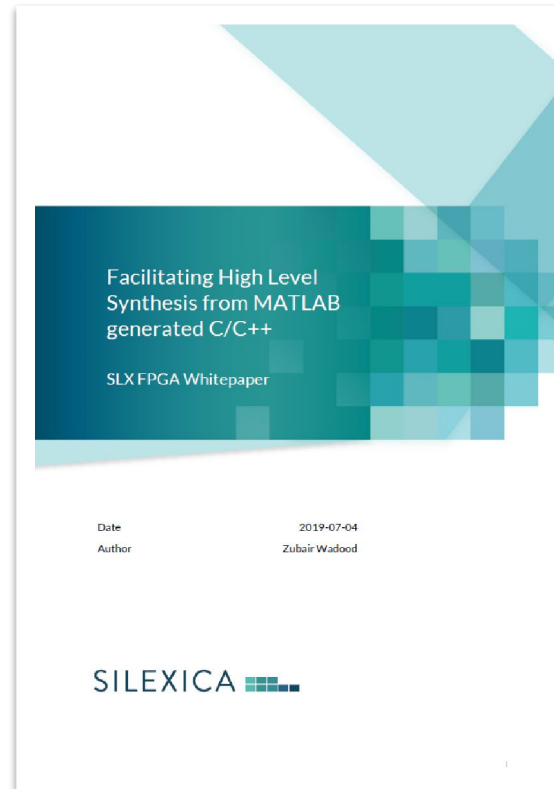
SLX FPGA Walkthrough Videos

SLX FPGA – Application White Papers



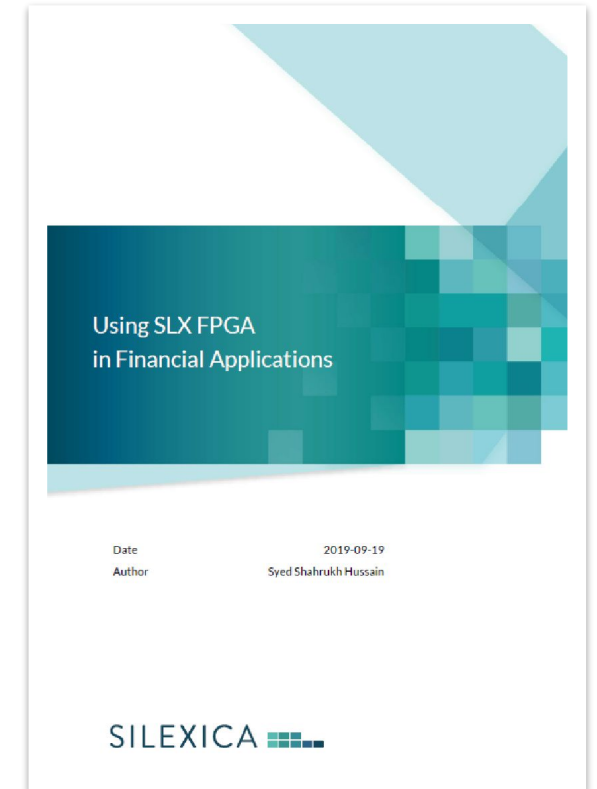
SHA-3 Algorithm

600x Speed-up with SLX FPGA



Kalman Filter

62x Speed-up with SLX FPGA



Black Sholes & Heston

29x Speed-up with SLX FPGA

Reduce Development Time with SLX FPGA

Design Phase	Done by hand (Days)	SLX Optimized (Days)
Clean up code to be synthesizable	1	0.5
Synthesize first HW	1	1
Refine Synthesis by inserting pragmas	10 days	1
Repeat last step until satisfied with results		2
Validate using C/C++ Testbench	0.5	0.5
Create IP	0.5	0.5
Total time	13*	6

- Example from expert HLS user
- Design was relatively small vision processing algorithm (~4KLUTs)
- Final performance achieved with HLS+SLX was better than RTL.

Adam Taylor Blog – Influential Xilinx Blogger

High Level Synthesis Made Easier!



Adam Taylor Following

Sep 13 · 4 min read

I have recently been evaluating the SLX FPGA tool from [Silexica](#). If you are not familiar with [SLX FPGA](#) it is designed to work with both Vivado HLS and SDSoc.



*“What is interesting to me having worked considerably with HLS over the years is how easy the insertion of pragmas was with SLX FPGA. HLS optimization can be a challenging, iterative and time-consuming process, **SLX FPGA made this much simpler.**”*

SLX Release Schedule

Release	Quarter	Date
SLX v20.1	Q1	April 6, 2020
SLX v20.2	Q2	June 29, 2020
SLX v20.3	Q3	September 21, 2020
SLX v20.4	Q4	December 14, 2020

Summary

- SLX FPGA accelerates the journey from C/C++ to Hardware by removing many of the roadblocks of using HLS
- SLX FPGA takes the guesswork out of using HLS
- SLX FPGA can help you maximize the performance of your designs in a fraction of the time



Silexica is ready to help you get started with SLX FPGA today!