Précis: A Design-Time Precision Analysis Tool

Mark L. Chang and Scott Hauck

Department of Electrical Engineering

University of Washington

Seattle, WA

Why precision analysis?

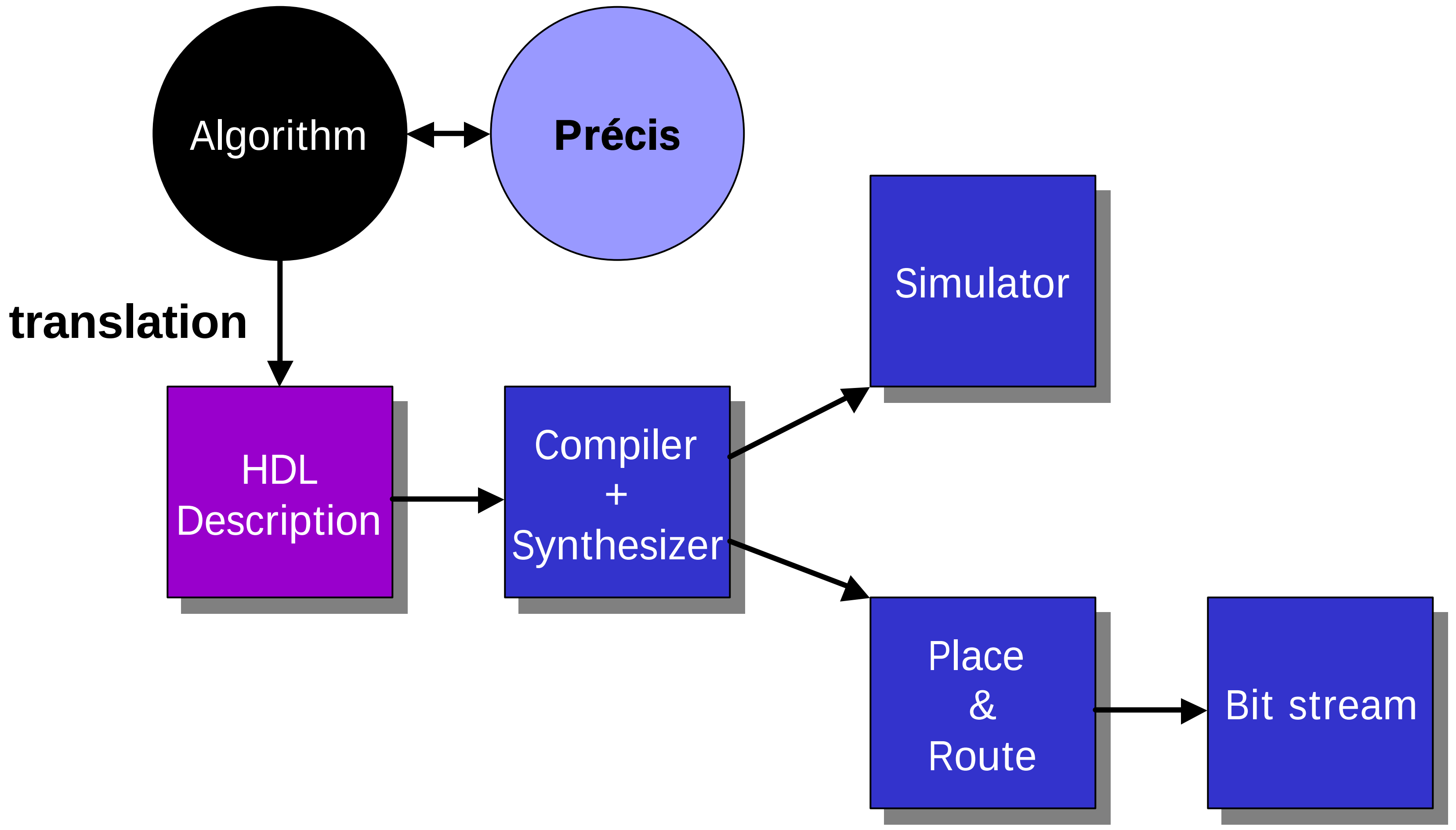
- We have an algorithm originally written for general purpose processors
- Want to implement it in an FPGA

- Software languages use data types to specify precision
  - Not all bits in the data types are necessarily used
- FPGAs can work at the single-bit level
  - Match data paths to the algorithm
  - Correctness
  - Optimality

Manual precision analysis

- What are the provable precision requirements of my algorithm?
- What are the actual precision requirements of my data sets?
- What are the effects of fixed-precision on my results?
- Where should I focus my efforts?

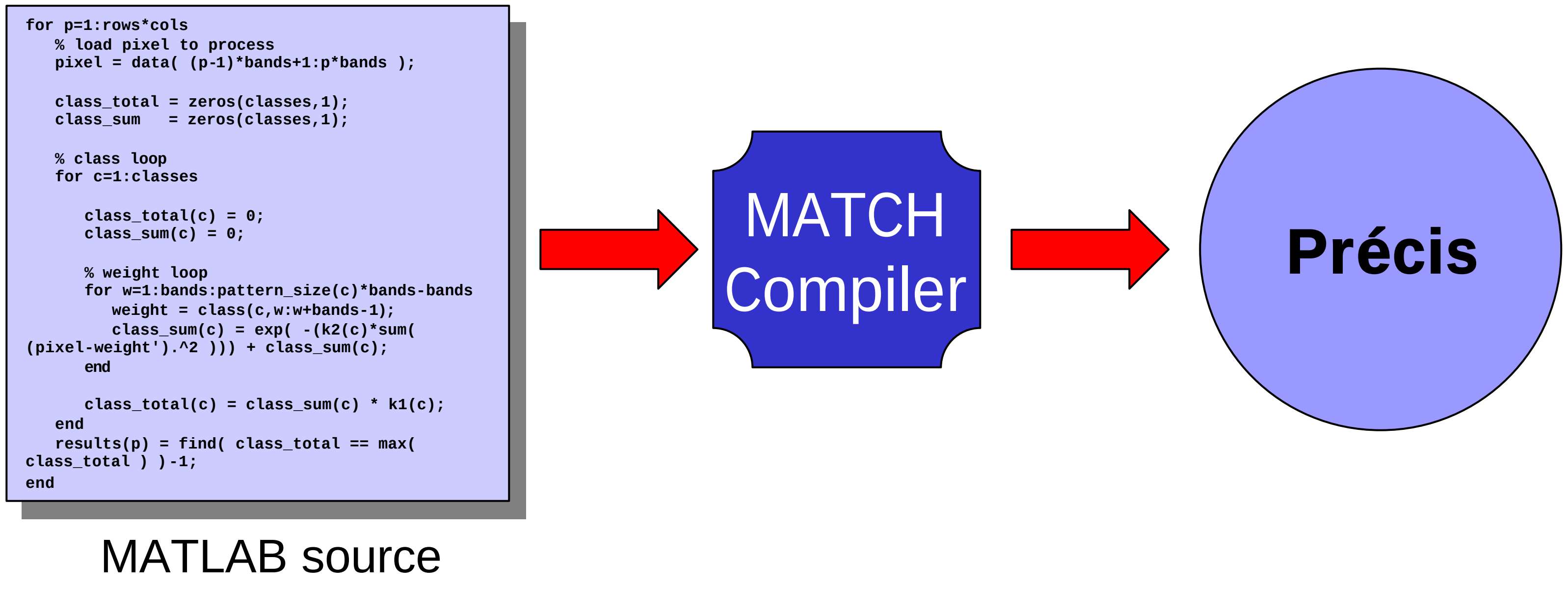
Role of Précis in Tool Chain



Précis

Compiler Front End

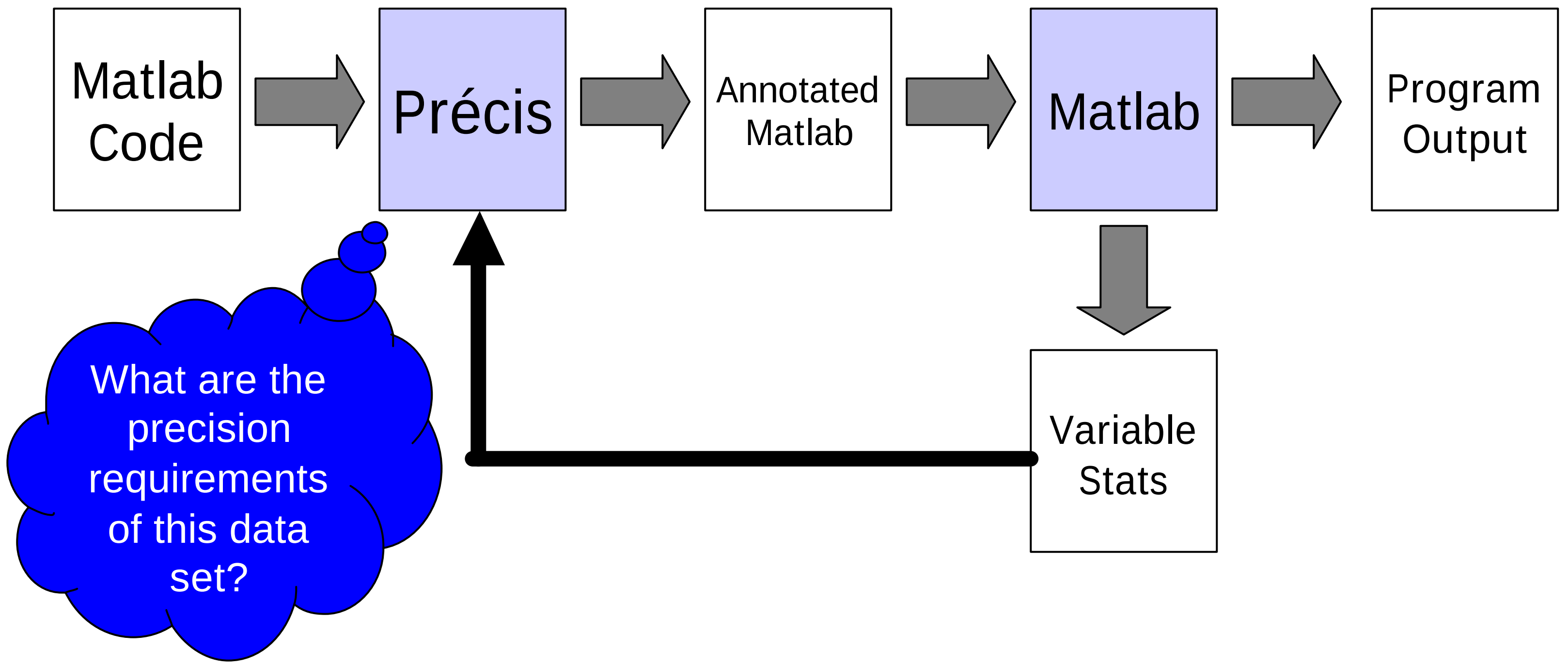
- MATCH compiler from Northwestern University
  - Understands a subset of the MATLAB grammar
  - Constructs an Abstract Syntax Tree based on the MATLAB source



```
for p=1:rows*cols
   % load pixel to process
   pixel = data( (p-1)*bands+1:p*bands );

   class_total = zeros(classes,1);
   class_sum   = zeros(classes,1);

   % class loop
   for c=1:classes

      class_total(c) = 0;
      class_sum(c) = 0;

      % weight loop
      for w=1:bands:pattern_size(c)*bands-bands
         weight = class(c,w:w+bands-1);
         class_sum(c) = exp( -(k2(c)*sum(
(pixel-weight').^2 ))) + class_sum(c);
      end

      class_total(c) = class_sum(c) * k1(c);
   end
   results(p) = find( class_total == max(
class_total ) )-1;
end
```

Range Finding & Profiling

- Generates annotated MATLAB
- Records ranges of variables for sample data sets
- Results are loaded into Précis to allow for more investigation



Range Finding Example

MATLAB Input

```
for x=a:b
 d = a*x+c;
```

Range Finding Output

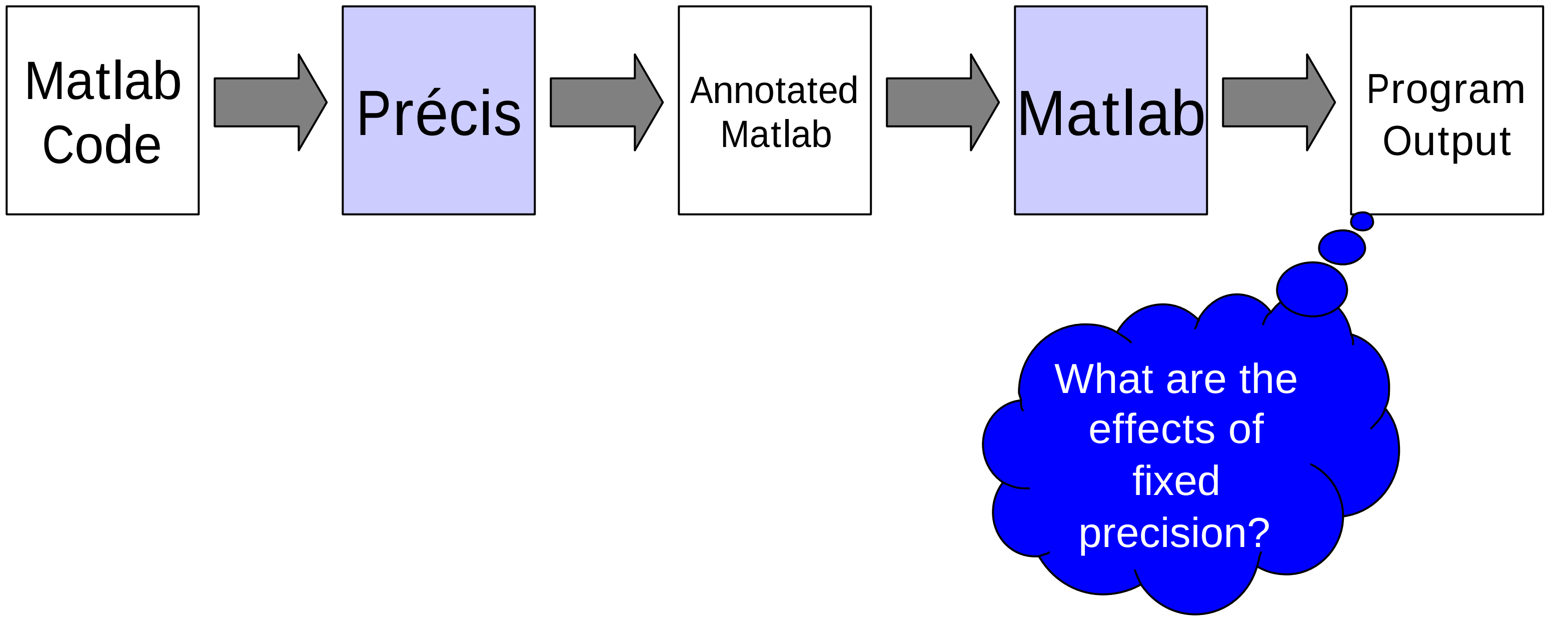
```
for x=a:b
 d=a*x+c;
 rangeFind(d,'rfv_d');
```

Example Results

```
rfv_d.min = 22
rfv_d.max = 1092
```

Simulation

- Generates annotated MATLAB
- User may specify precision constraints on any variable
- Simulation demonstrates the effects of fixed-point operations which may result in rounding or truncation errors in the output



Simulation Example

MATLAB Input

```
a = 128;
b = 2098;
c = 33276;
d = (a+(b*c));
e = a * d;
f = d + e;
```

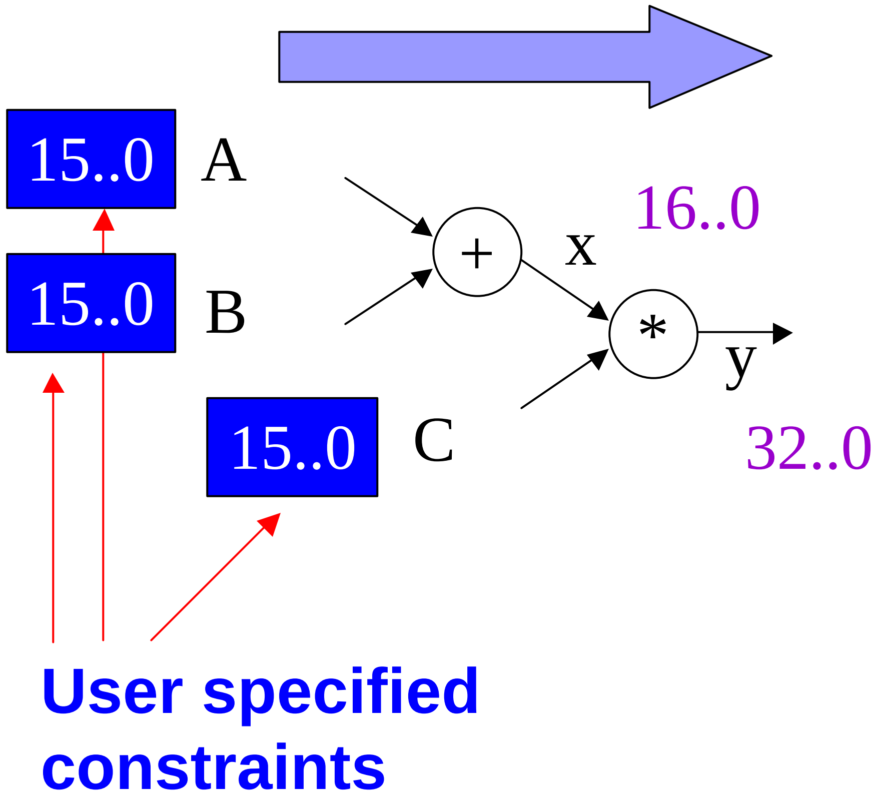
Constrain to $[2^{12}..2^{0}]$ and utilize truncation on both MSB and LSB.

Annotated MATLAB Output

```
 . . .
e=a*(fixp(d,12,0,’trunc','trunc’);
f=fixp(d,12,0,’trunc’,’trunc’)+e;
```

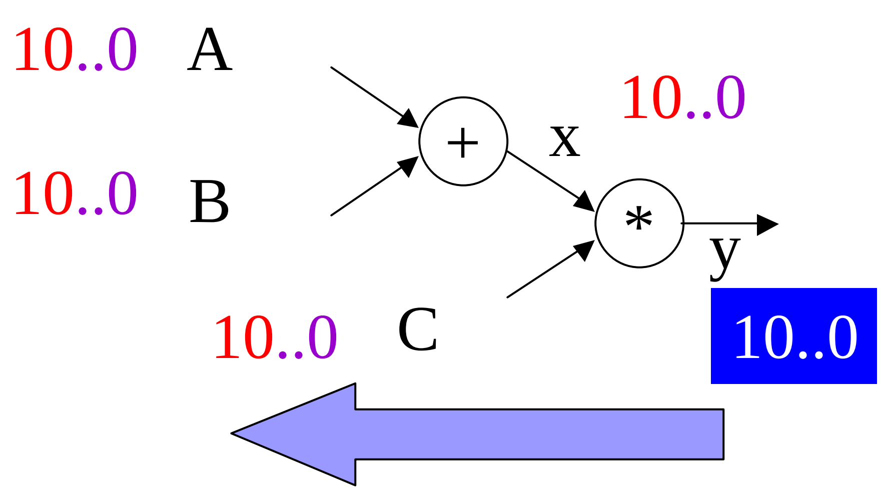
Propagation Engine

- What are the provable precision requirements of my algorithm?



```
x = a + b
y = c * x
```

Propagation Engine



```
x = a + b
y = c * x
```

- **User constrains output “y” to 10..0**
- **Reverse propagate**

Slack Analysis

- Where should I focus my optimization efforts?

- Propagation == upper bound
- Range finding == lower bound
- *Difference == Slack*

- Try to identify nodes that have the greatest area impact on the final circuit implementation

Slack Analysis

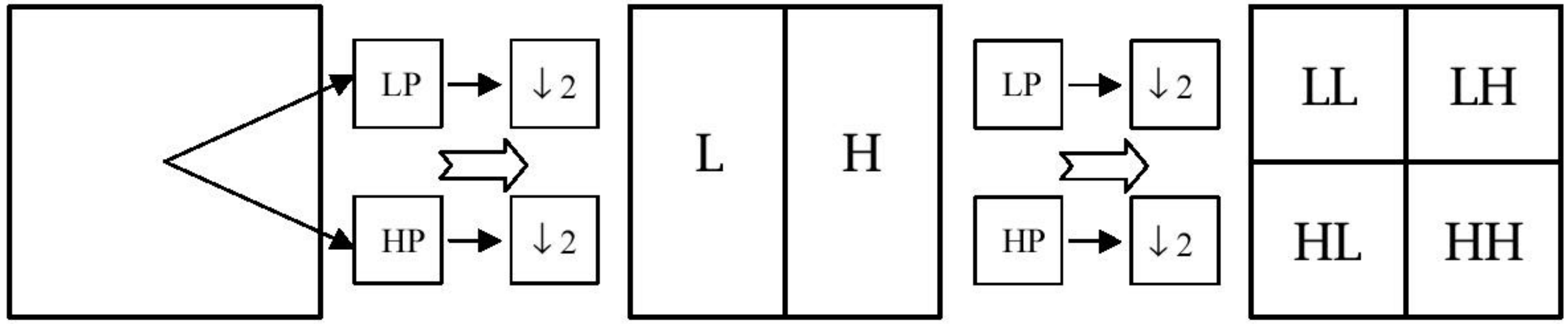
- For each node with slack, set precision to lower bound
  - Propagate change through system
  - Calculate the gain in area for this “move”
  - Area estimation determined from operator models
- This creates a “tuning list” of variables to consider
- User can choose to make moves and recalculate what the next move should be

User Guidance

- Want to guide a developer’s manual optimization
  - Helpful for a novice designer
  - Provides a starting point for hand-optimization
  - Allows iterative optimization of the implementation
- We ask questions, developer answers
  - What is the algorithm
  - Known precision of variables
  - Simulation and data gathering
- Provide suggestions to the user

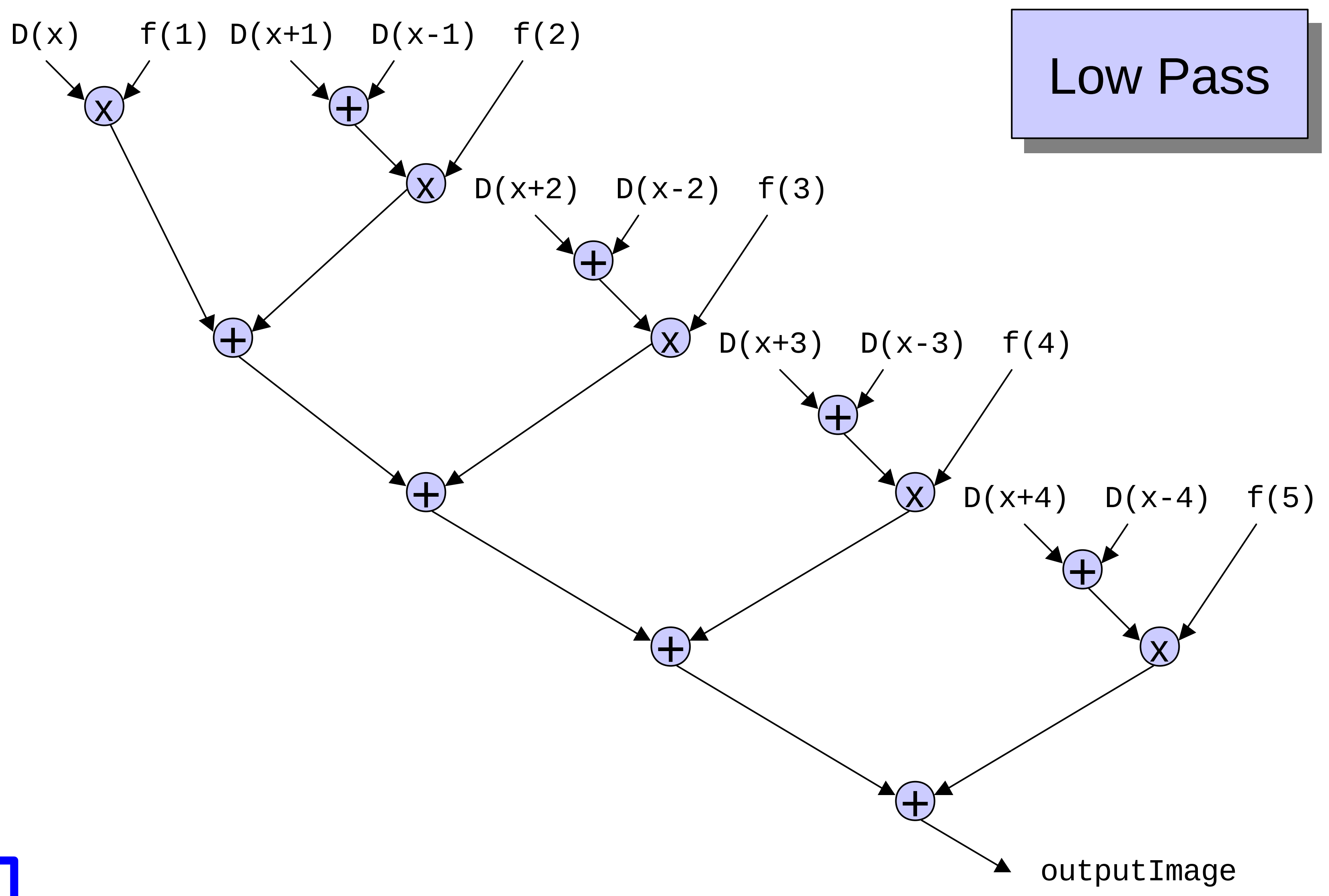
Benchmark: Wavelet Transform

- Perform high-pass and low-pass filtering in each dimension
- Follow with down-sampling to produce two half-sized sub-band images
- Repeat for as many levels as desired on LL sub-band

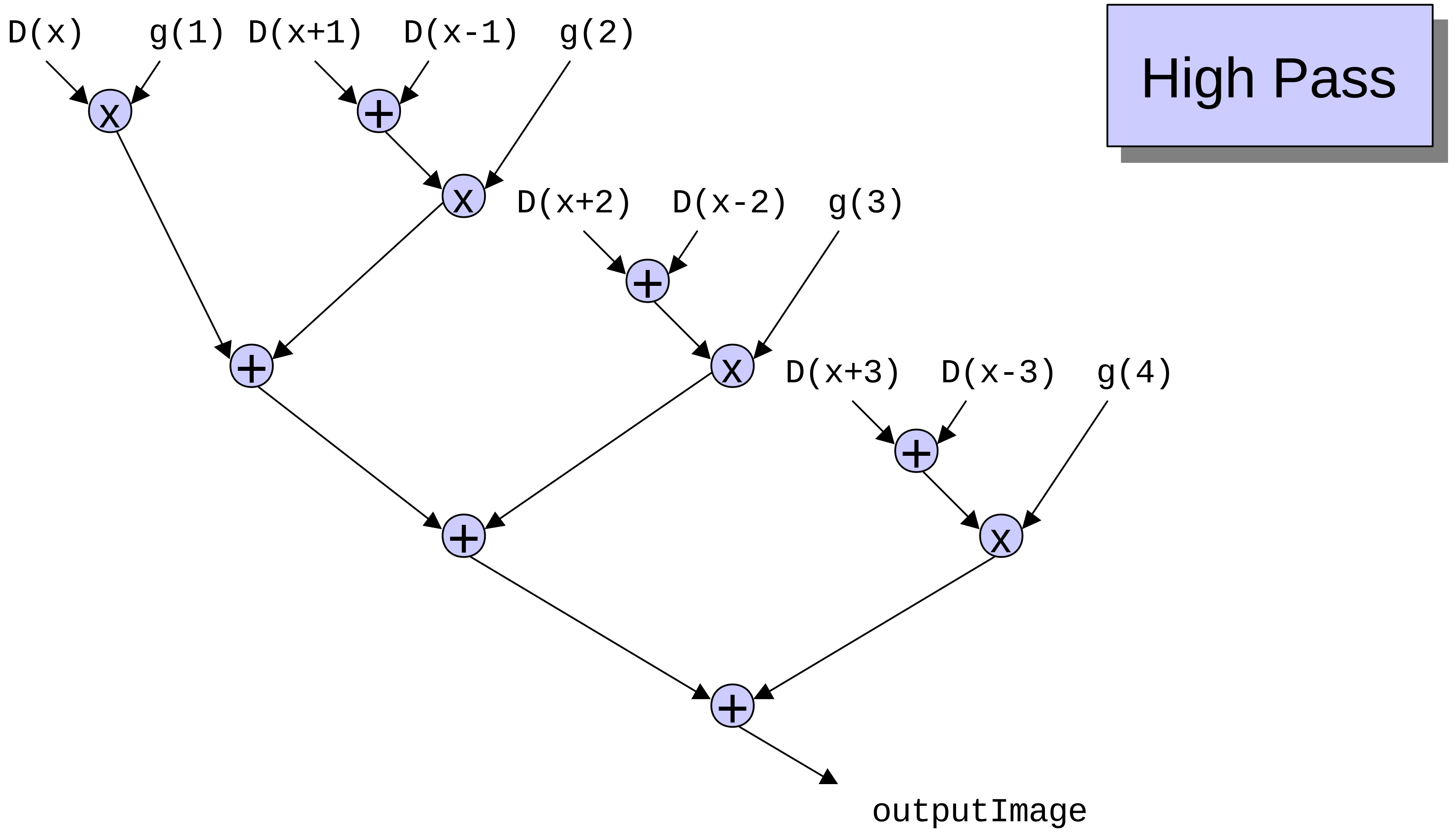


a) Original Image b) Horizontal Pass b) Vertical Pass

Wavelet Transform Structure

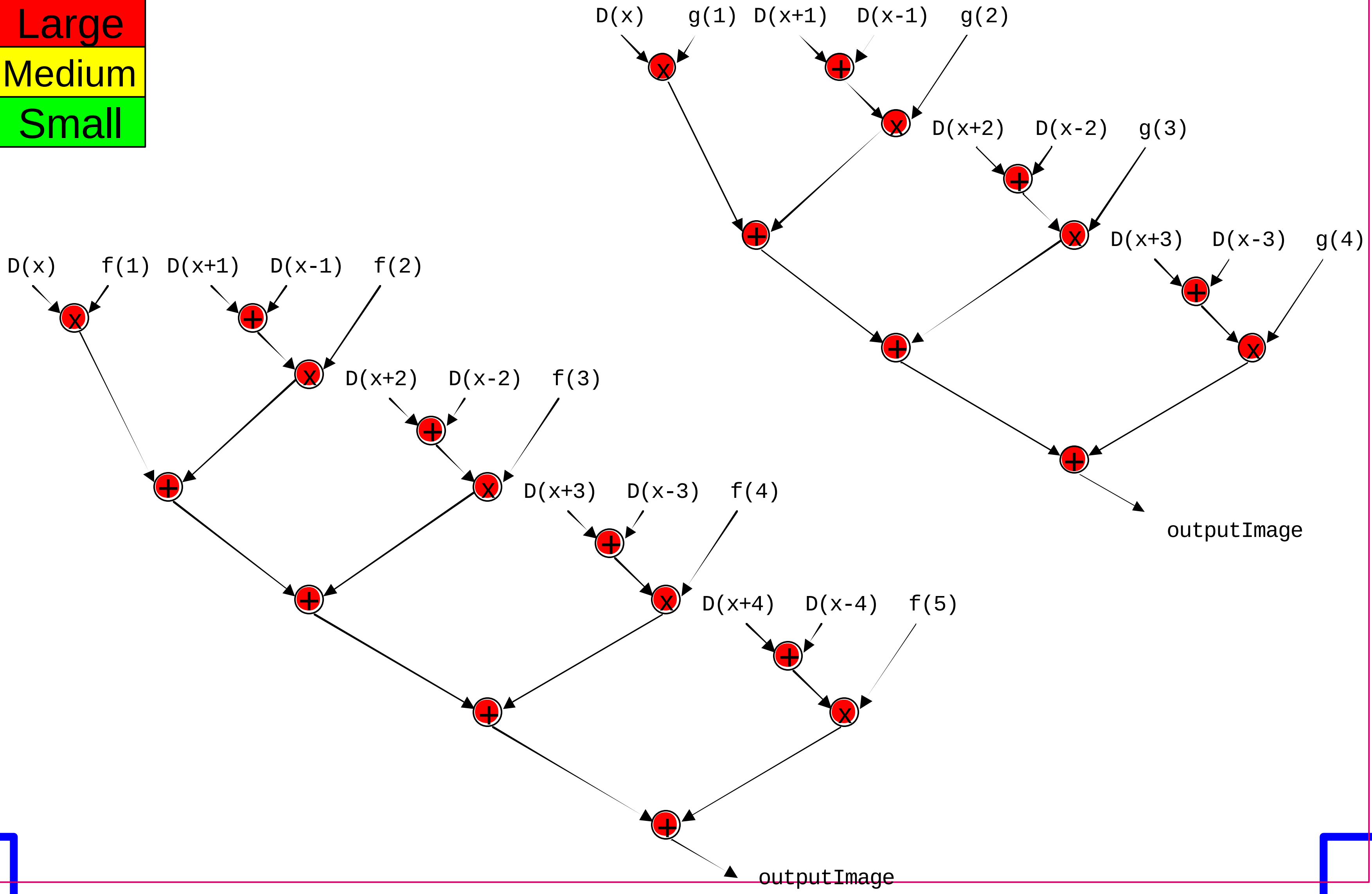


Wavelet Transform Structure



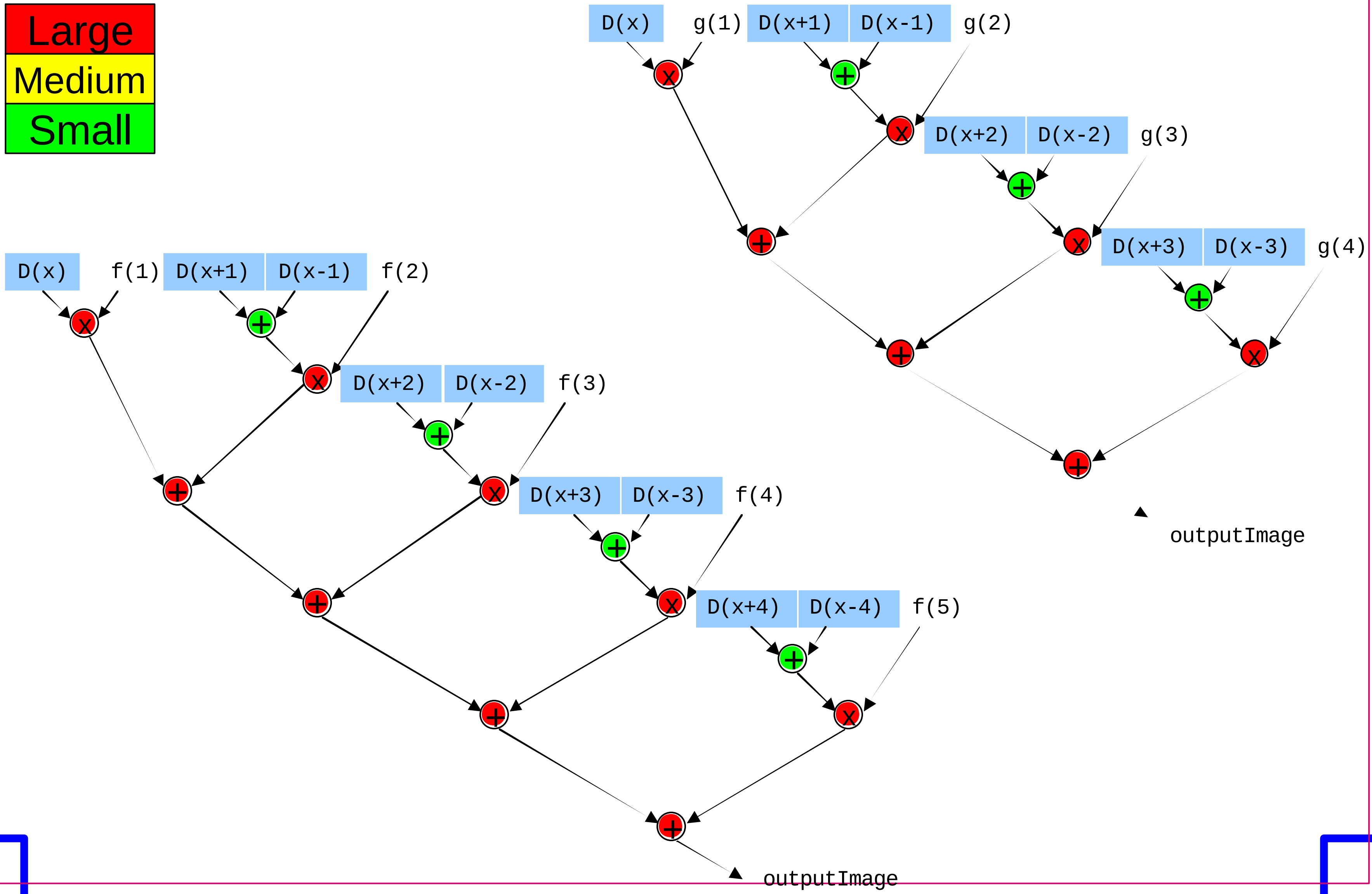
Wavelet Transform Structure

🟥 Large
🟨 Medium
🟩 Small



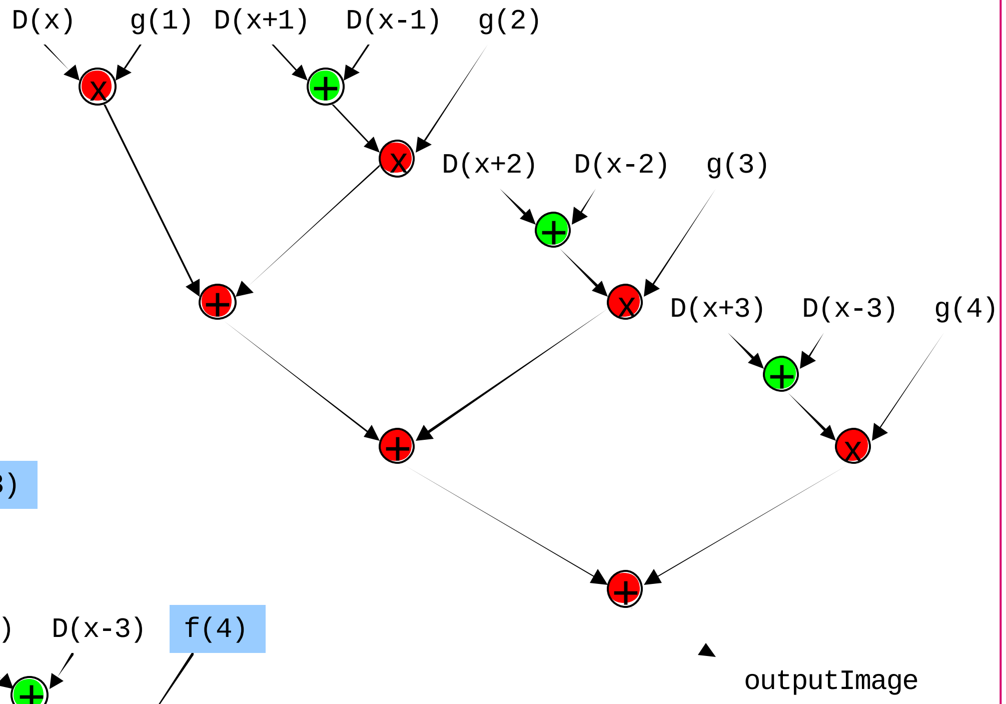
Wavelet Transform Structure

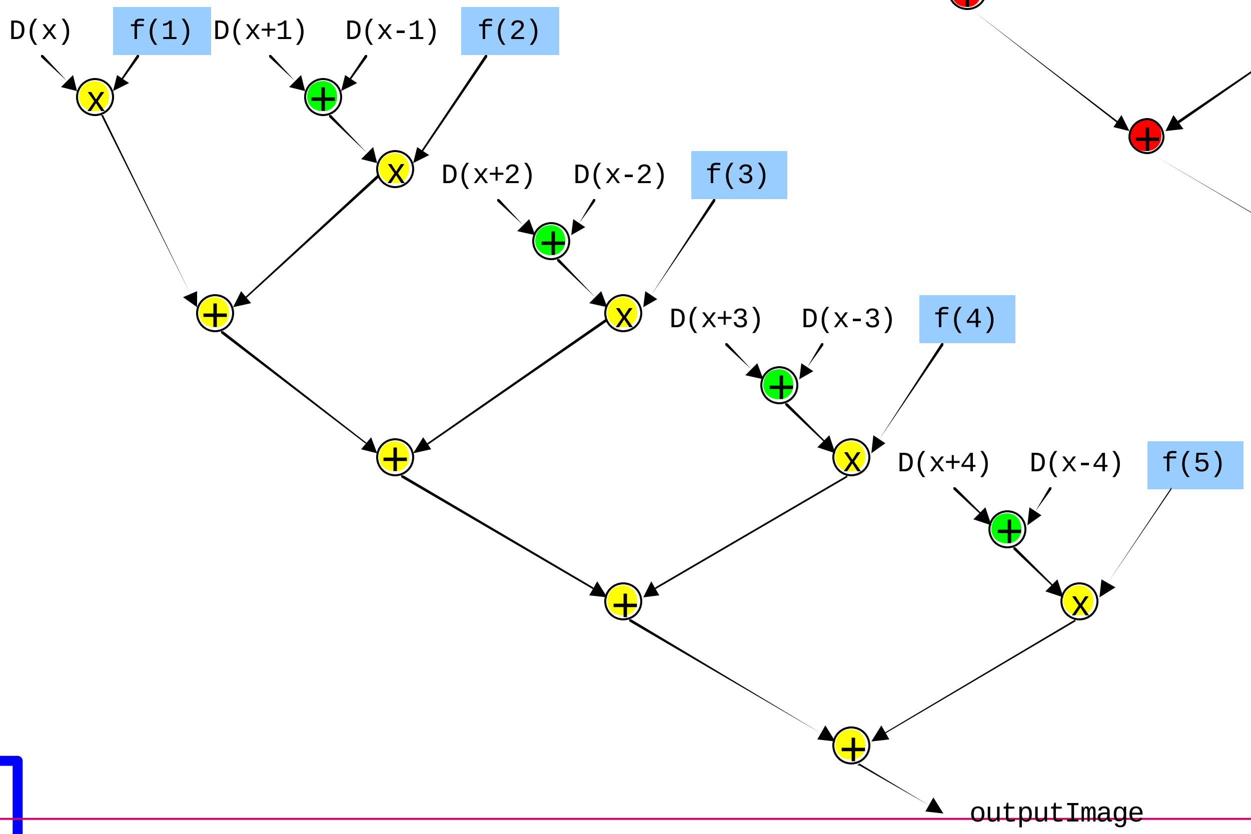
🟥 Large
🟨 Medium
🟩 Small



Wavelet Transform Structure

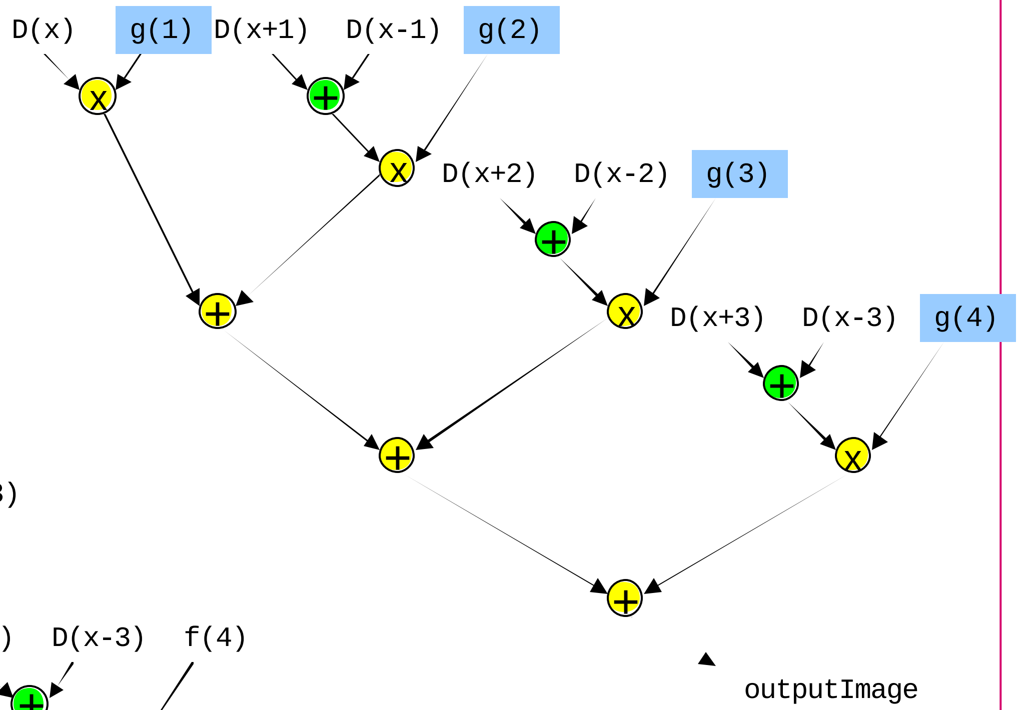
🟥 Large
🟨 Medium
🟩 Small

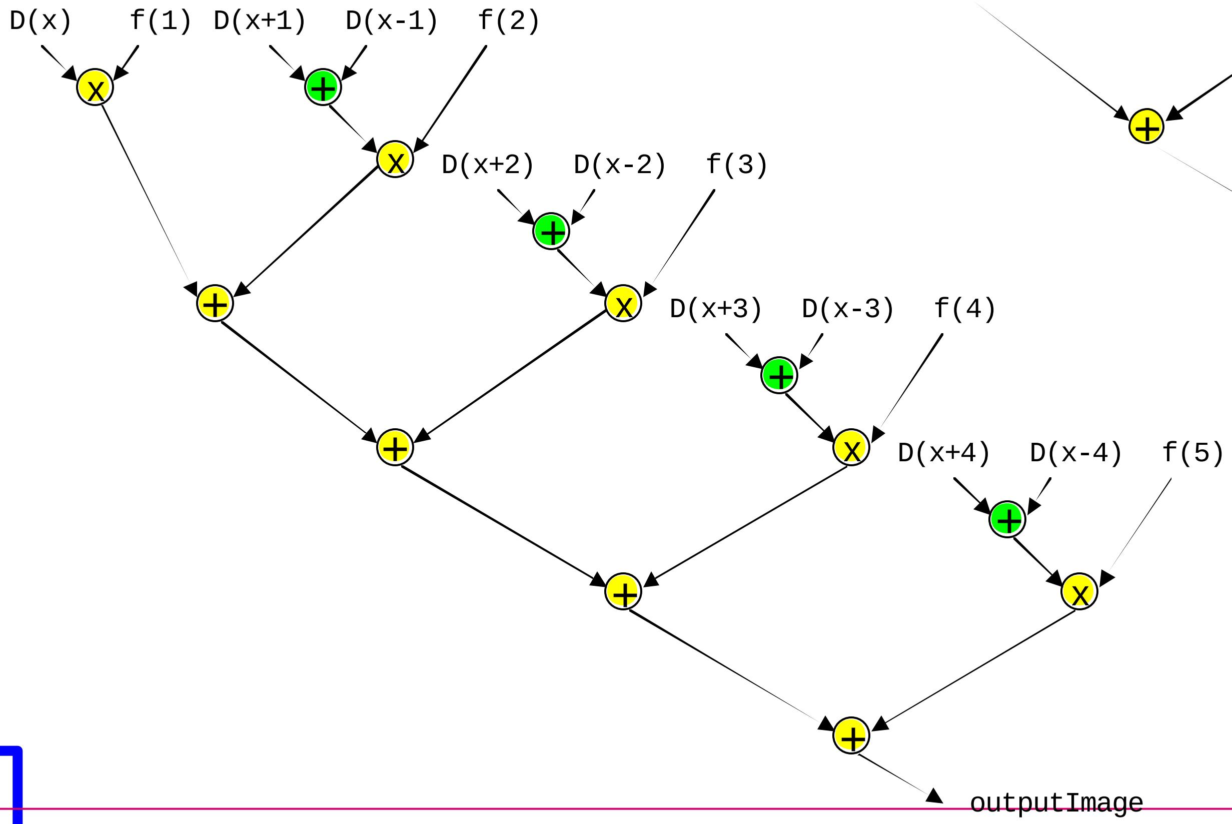




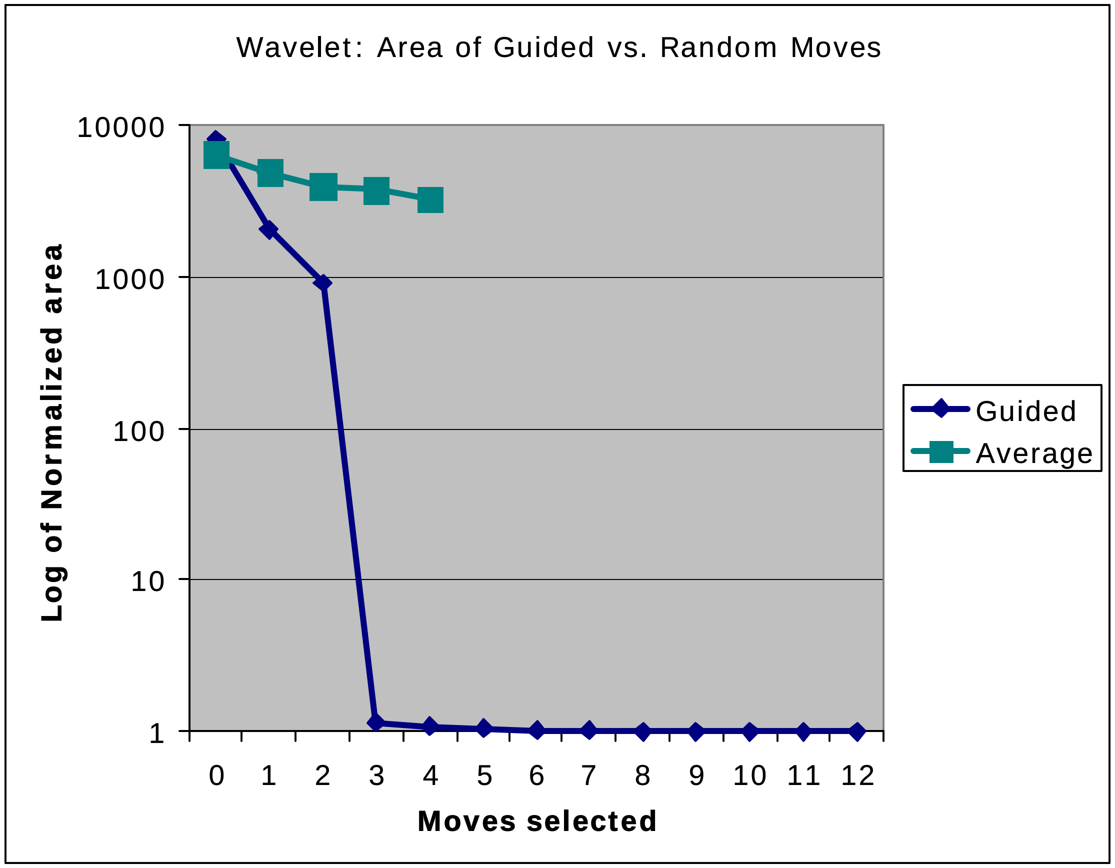
Wavelet Transform Structure

🟥 Large
🟨 Medium
🟩 Small



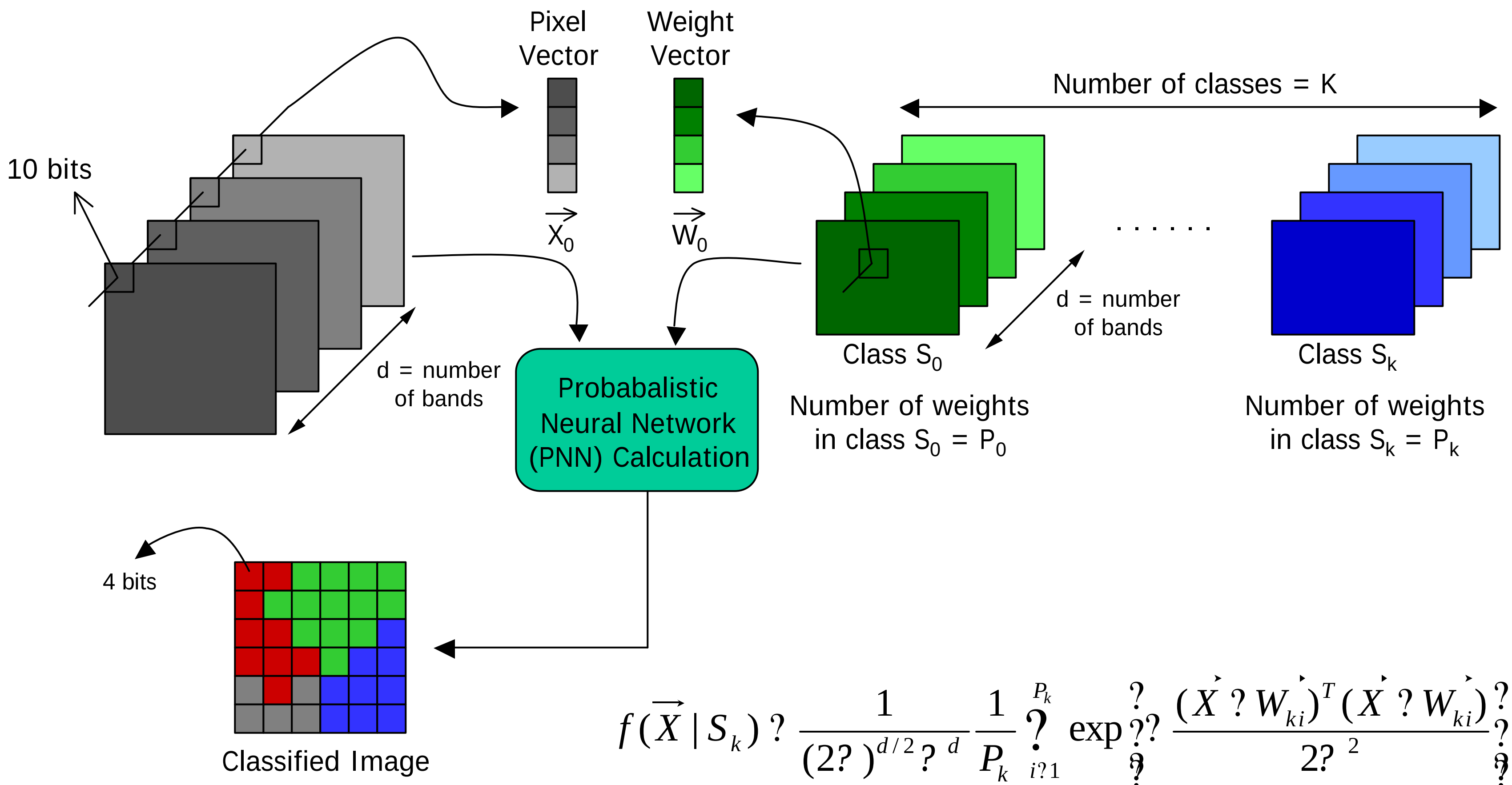


Wavelet Transform



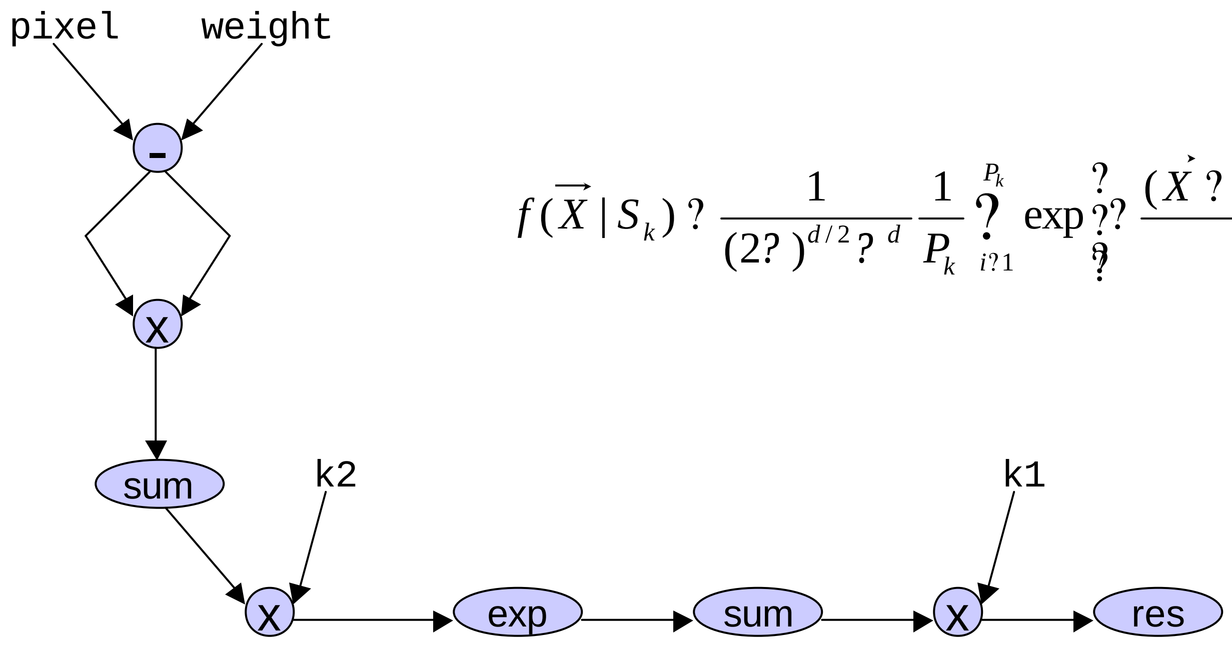
- 27 variables selected for slack analysis
- 3 moves to within 15% of lower bound
- 4 moves to within 10% of lower bound
- 7 moves to within 3% of lower bound

Benchmark: PNN



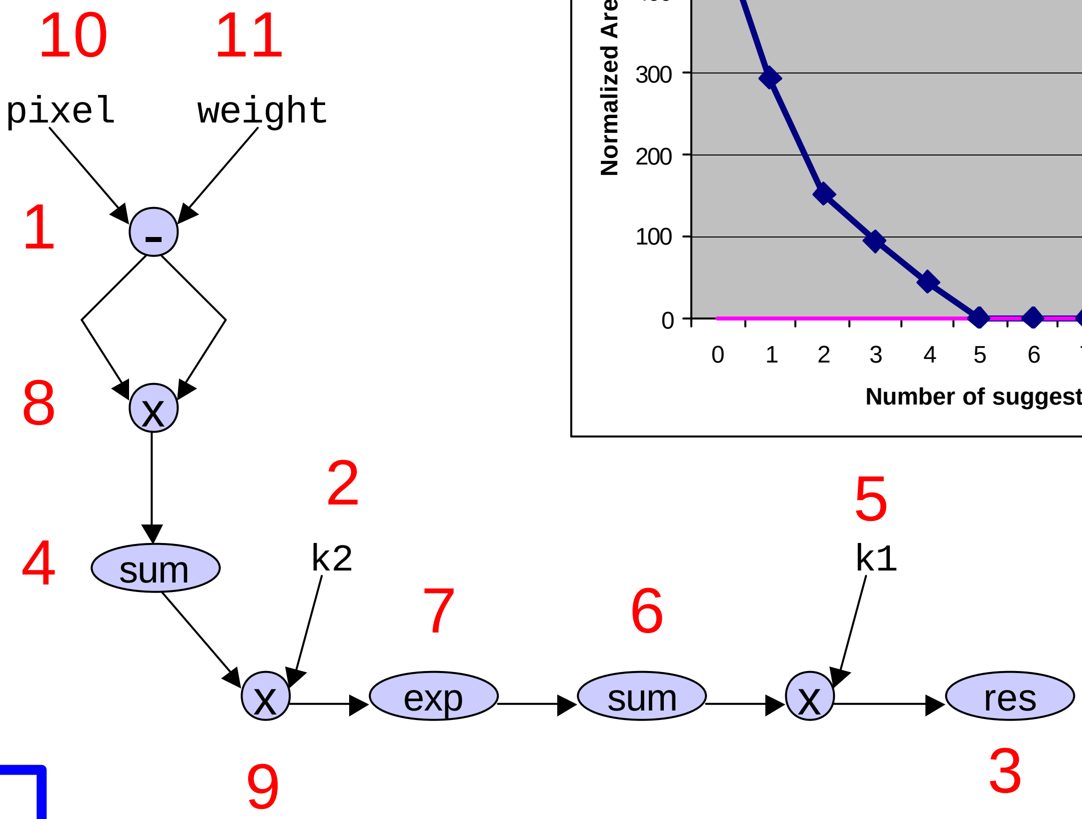
$$f(\vec{X} \mid S_k)\ ?\ \frac{1}{(2?)^{d/2}\ ?^{\ d}}\ \frac{1}{P_k} \overset{P_k}{\underset{i?1}{?}} \exp ?? \frac{(\vec{X}\ ?\ \vec{W}_{ki})^T (\vec{X}\ ?\ \vec{W}_{ki})}{2?^{\ 2}} ?$$

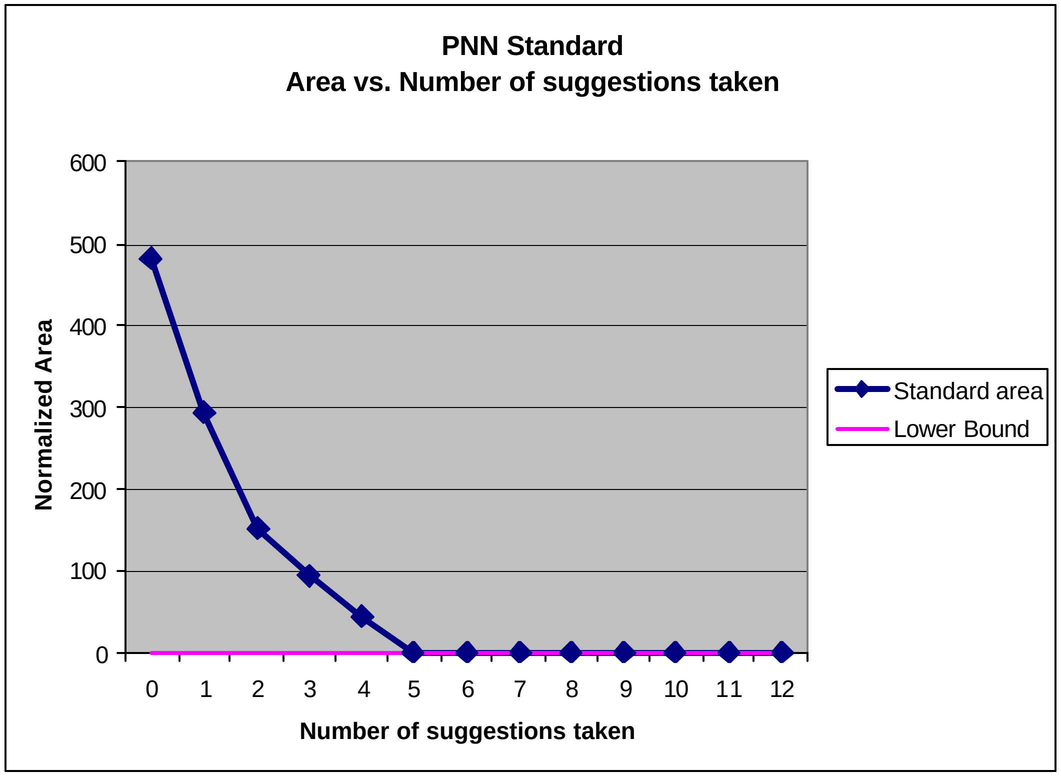
PNN: Structure



$$f(\overrightarrow{X} \mid S_k) ? \frac{1}{(2?)^{d/2} ?^{d}} \frac{1}{P_k} \overset{P_k}{\underset{i?1}{?}} \exp ??? \frac{(\vec{X} ? \vec{W}_{ki})^T (\vec{X} ? \vec{W}_{ki})}{2?^{2}} ???$$

PNN: Moves suggested



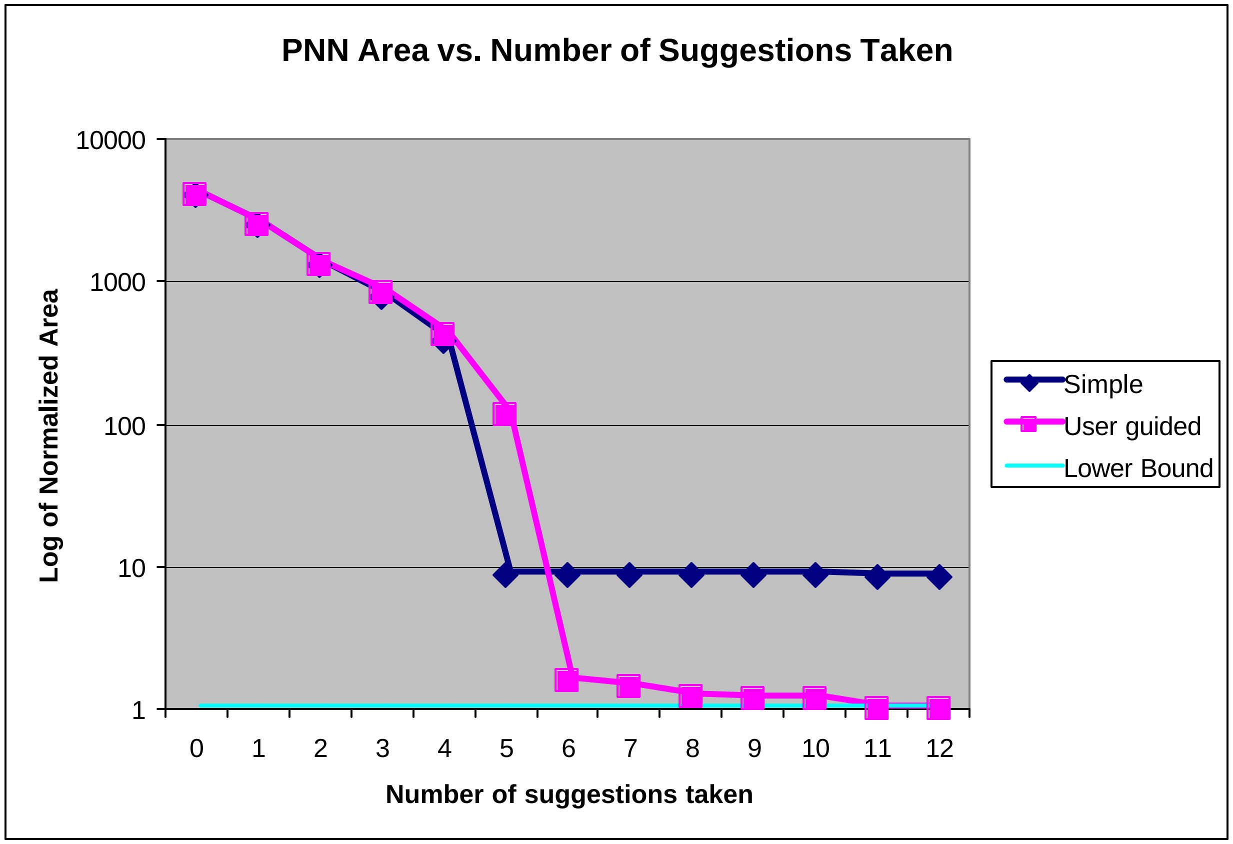


- 4 moves to within 50x
- 5 moves to within 5%
- 8 moves to within 2.5%

PNN: Experience Counts

- Ranges discovered by range finding phase may be too “wide”
  - Values that are very small and near zero require many bits of precision to the right of the decimal point
  - Automated range-finding phase cannot determine at what point values become too small to be significant
- User can re-constrain variables to more sensible values
- Utilize simulation to determine how much error is tolerable

PNN: Re-Constrained Results



- Simple reaches within 10x in 5 moves
- Guided method achieves 8x better lower bound than simple

Conclusions

- Introduced a software tool for interactive precision analysis at design time
- Simplifies typical precision analysis tasks
  - Simulation, range finding, constraint propagation
- Provides an effective methodology for suggesting the order of optimization steps