MICROELECTRONIC SYSTEMS DESIGN RESEARCH GROUP

CHPC

GRK 1932

Near Real-Time Risk Simulation of Complex Portfolios on Heterogeneous Computing Systems with OpenCL

Javier Alejandro Varela, Norbert Wehn
Microelectronic Systems Design Research Group
University of Kaiserslautern, Germany

Javier Alejandro Varela, Norbert Wehn. 2017. Near Real-Time Risk Simulation of Complex Portfolios on Heterogeneous Computing Systems with OpenCL. In Proceedings of IWOCL '17, Toronto, Canada, May 16-18, 2017, 10 pages.
DOI: http://dx.doi.org/10.1145/3078155.3078161

Javier Varela

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Summary

I. Overview: Portfolio Risk

II. OpenCL + Workload Allocation

III. Algorithmic Optimizations / Numerical Scheme

IV. Minimizing Device Global Memory

V. Conclusion

MICROELECTRONIC SYSTEMS DESIGN RESEARCH GROUP

Summary

I. Overview: Portfolio Risk

II. OpenCL + Workload Allocation

III. Algorithmic Optimizations / Numerical Scheme

IV. Minimizing Device Global Memory

V. Conclusion

Overview: Portfolio Risk

Portfolio Composition:

- **Stocks** (S_1, S_2)
- **Corporate Bond**
- **Foreign Currency**
- **Options:**
  - European Asian (on S_1)
  - European Barrier (on S_2)
  - American Vanilla (on S_1)
  - American 2D Max (on S_1,S_2)
- **Cash** (no simulation required)

What is the **RISK** at which this portfolio is exposed ?

Overview: Portfolio Risk

What is Value-at-Risk (VaR) ?

*„I am **α** percent certain there will not be a loss of more than **VaR** USD in the next **N** days"* (Hull, *Options Futures and Other Derivatives*).

VaR of level **α**

Portfolio loss function L

α - quantile

$$Var_\alpha(\tilde{L}_{t+N}) \triangleq \inf_{l \in R}\{P(\tilde{L}_{t+N} > l) \leq (1-\alpha)\} = \inf_{l \in R}\{\tilde{F}_L(l) \geq \alpha\}$$

Loss empirical distribution function

$$\tilde{F}_L(l) \triangleq P(\tilde{L}_{t+N} > l)$$

where α is as high as 95%, 99%, 99.5%.

Industry standard (Basel II and Basel III)

What is Expected Shortfall (ES or cVaR) ?

$$ES_\alpha(\tilde{L}_{t+N}) \triangleq \frac{1}{1-\alpha}\int_\alpha^1 Var_\gamma(\tilde{L}_{t+N})\, d\gamma$$

Overview: Portfolio Risk

How to compute Value-at-Risk (VaR) ?

1. ***Historical simulation:*** describes future changes …
   based on empirical distribution of *observed past data*.
2. ***Variance-Covariance method:*** first-order *approximation* of the loss function L.
   assumes normally distributed returns.
3. ***Monte Carlo (MC) method***: *simulates* the loss function L.

- ✓ Does not rely on historical data.
- ✓ No need for approximation.
- ✓ No assumption of a normal distribution.
- ❖ Computationally intensive problem
- ❖ Heterogeneous problem (many different algorithms involved)

Exploit OpenCL

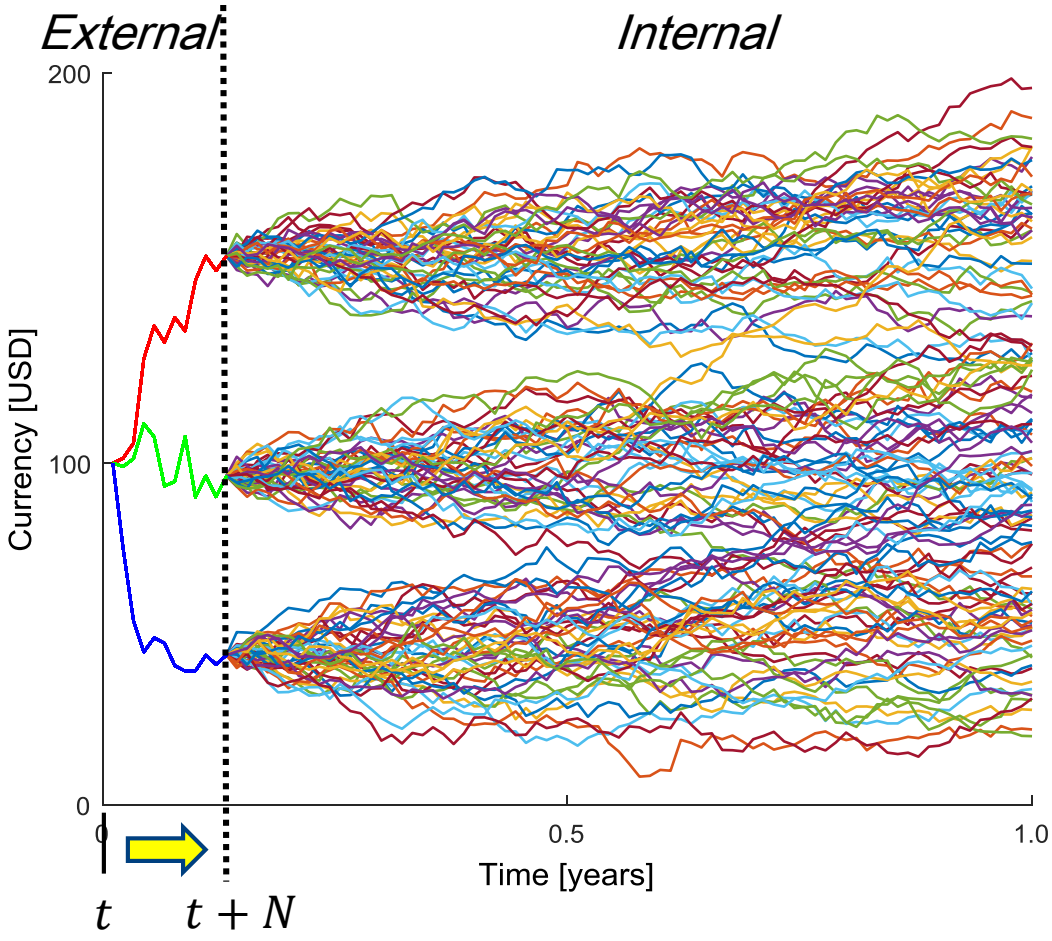
Overview: Portfolio Risk

Portfolio Simulation:

- **Stocks** (S_1, S_2)
- **Corporate Bond**
- **Foreign Currency**

- **Options:**
  - European Asian (on S_1)
  - European Barrier (on S_2)
  - American Vanilla (on S_1)
  - American 2D Max (on S_1,S_2)

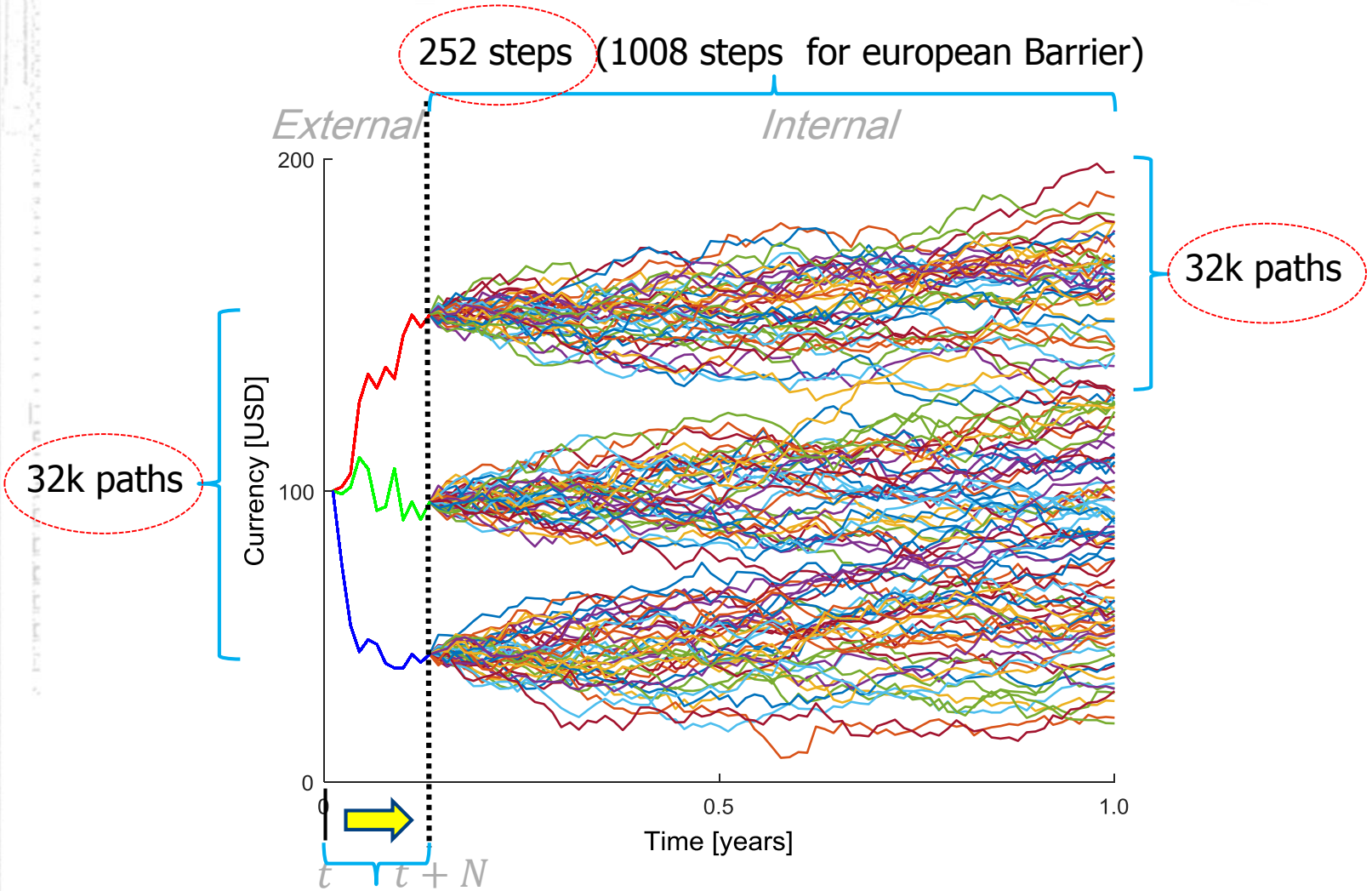
- **Cash** (no simulation required)



Two Models:

- Black-Scholes (BS)
- Heston

Overview: Portfolio Risk

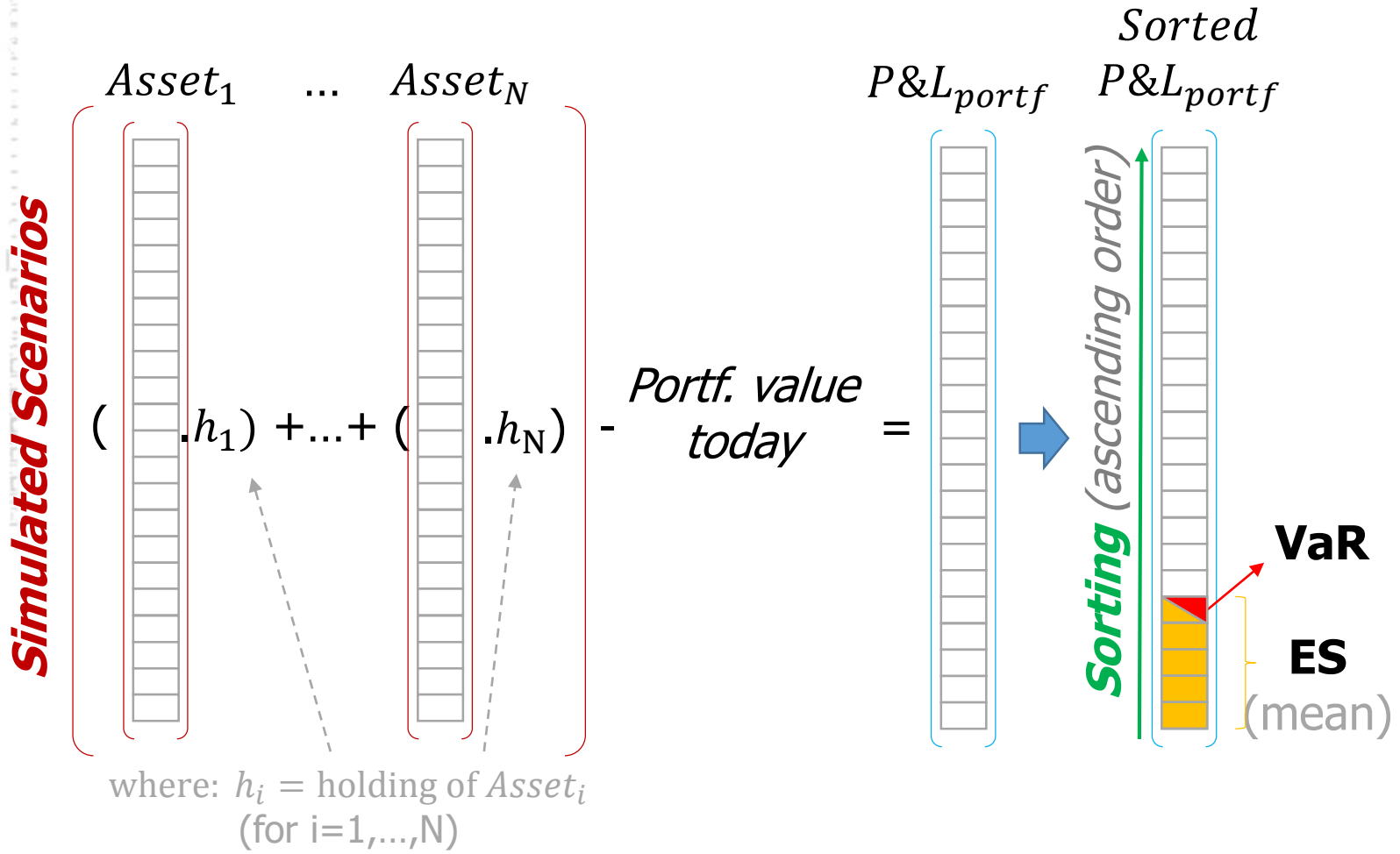


1 step (BS model)
4 steps (Heston model)

Single-precision floating point operations

Overview: Portfolio Risk

Sequence of steps:



MICROELECTRONIC SYSTEMS DESIGN RESEARCH GROUP

Summary

I. Overview: Portfolio Risk

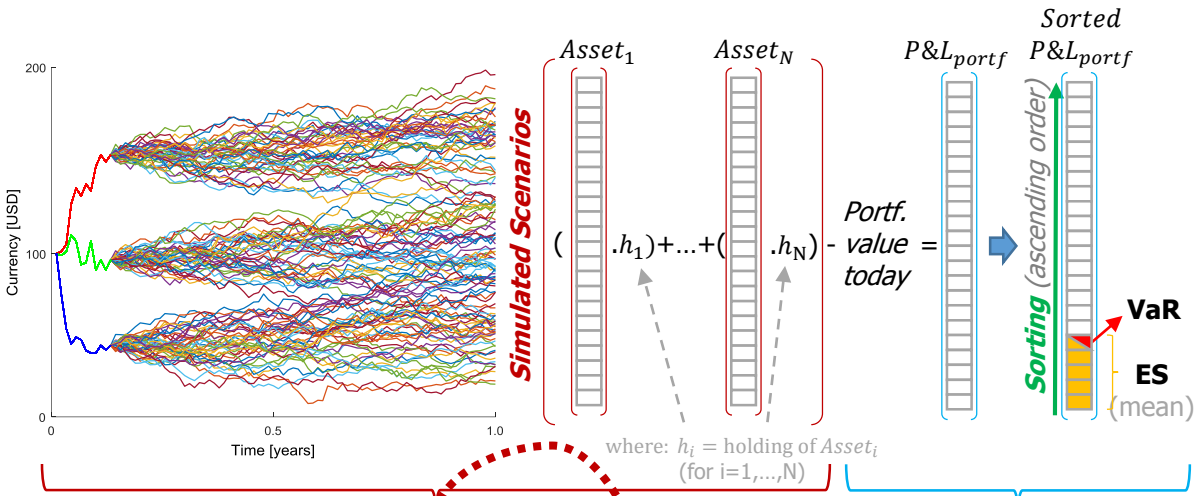
II. OpenCL + Workload Allocation

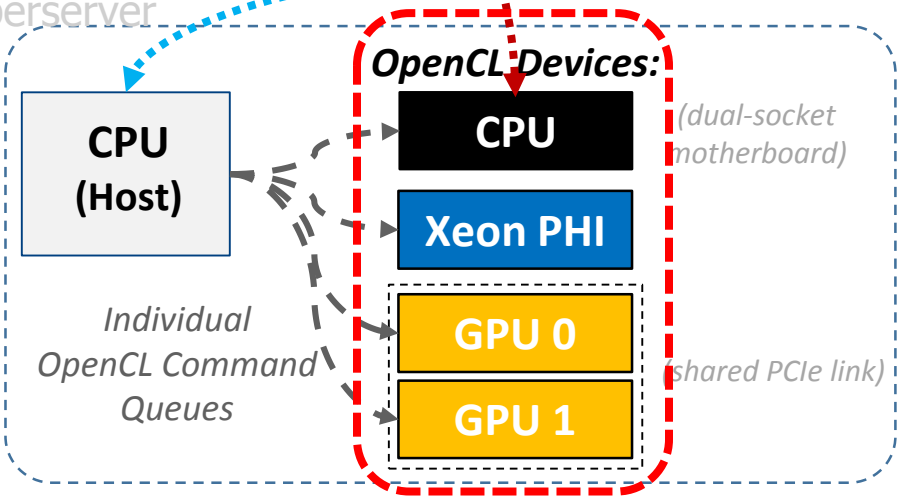
III. Algorithmic Optimizations / Numerical Scheme

IV. Minimizing Device Global Memory

V. Conclusion

OpenCL + Workload Allocation





SuperMicro Superserver 7048GR-TR

Intel Xeon E5-2670 v3

Intel Xeon E5-2670 v3

Intel Xeon Phi 7120P

Nvidia K80

OpenCL + Workload Allocation

ILP formulation

(Matlab's *intlinprog*)

$$\min_x f^T.x \; subject\; to \begin{cases} x_i \; are \; integers \\ A.x \leq b \\ A_{eq}.x \leq b_{eq} \\ lb \leq x \leq lu \end{cases}$$

Set of Kernels: $K = \{k_{Stocks2Corr}, k_{Bond}, k_{F.Currency}, \cdots\}$

Set of Devices: $D = \{d_{CPU}, d_{PHI}, d_{GPU0}, \cdots\}$

Binary decision variables: $x_{k,d}$, where: $0 \leq x_{k,d} \leq 1$

Total runtime: x_T , where: $0 \leq x_T \leq \infty$

OpenCL + Workload Allocation

ILP formulation … (continued)

Constraints:

Single assignment of each kernel among all devices:

$$\forall k \in K, \sum_{d \in D} x_{k,d} = 1$$

Device runtime:

$$\forall d \in D, \sum_{k \in K} t_{k,d} . x_{k,d} - x_T \leq 0$$

Device global memory:

$$\forall d \in D, \sum_{k \in K} m_{k,d} . x_{k,d} \leq M_d$$

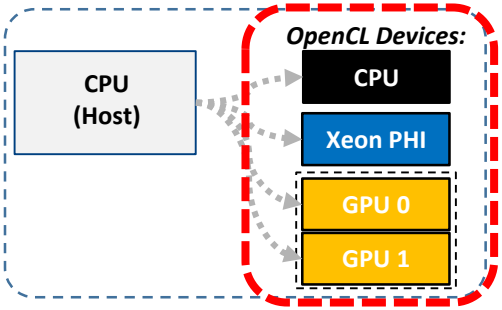
Cost function:

$$f^T . x = x_T$$

OpenCL + Workload Allocation

Workload Allocation with ILP

1) Profile each kernel on every device → 2) ILP →

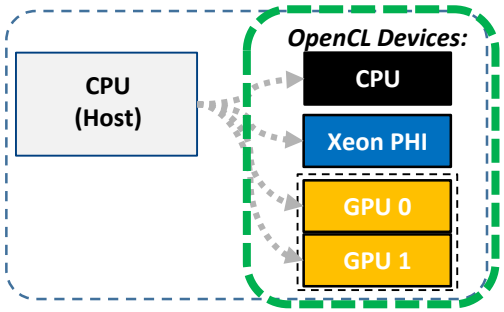


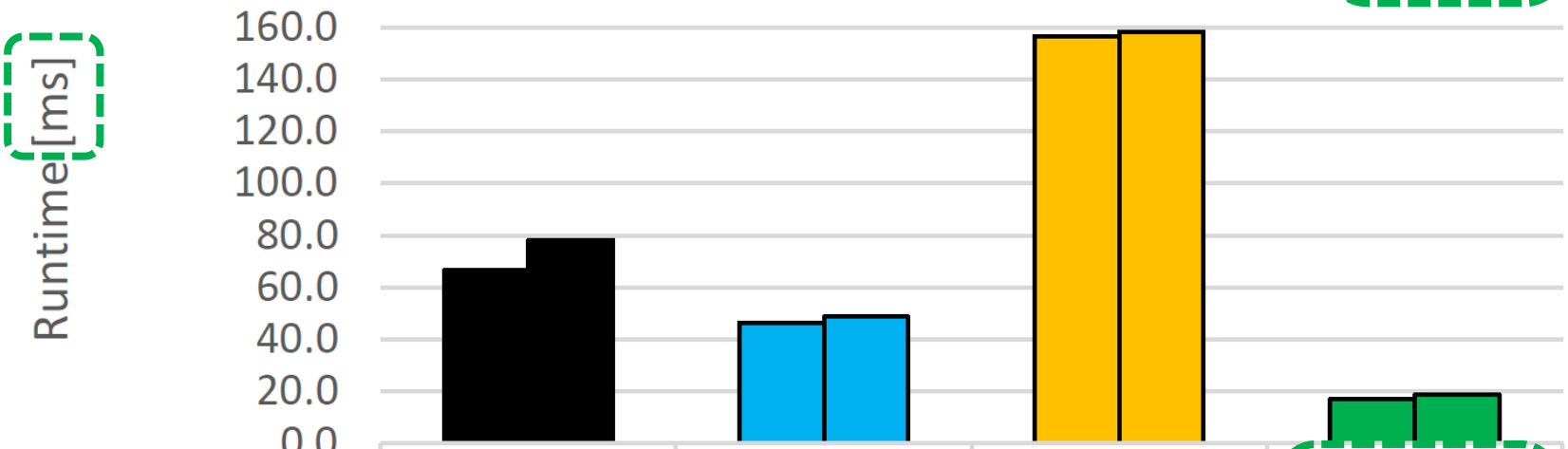
Kernel	Model	Runtime CPU	PHI	GPU
External Simulation				
Stocks2Corr	BS	0.12 ms	0.28 ms	0.02 ms
	Heston	0.15 ms	0.43 ms	0.03 ms
Bond	-	0.22 ms	0.37 ms	0.05 ms
F.Currency	-	0.04 ms	0.30 ms	0.01 ms
Internal Simulation: European Options				
Asian(RNs)	BS	8.89 s	6.59 s	6.74 s
	Heston	17.39 s	11.43 s	11.09 s
Asian(InfoP1)	BS	2.83 ms	1.14 ms	0.28 ms
2DMax(RNs)	Heston	399.24 s	572.37 s	91.92 s
2DMax(PathsP1)	BS	14.31 ms	13.78 ms	0.99 ms
	Heston	19.02 ms	20.52 ms	1.45 ms
2DMax(PathsP2)	-	65.33 s	47.78 s	41.20 s
– SurfGen	-	78.89 s	54.93 s	48.11 s
2DMax.Interp	-	0.05 ms	0.32 ms	0.01 ms

Kernel	CPU	PHI	GPU0	GPU1
Model: BS				
Stocks2Corr	–	–	–	√
Bond	–	–	–	√
F. Currency	–	–	–	√
Euro. Asian (*Info-reuse*)[1]	–	–	√	–
Euro. Barrier (*Info-reuse*)[1]	–	√	–	–
Amer. Vanilla. BT	√	–	–	–
Amer.2DMax (*Interp+S.Load*)	–	–	–	√
Model: Heston				
Stocks2Corr	√	–	–	–
Bond	√	–	–	–
F. Currency	–	–	√	–
Euro. Asian (*Info-reuse*)[1]	–	–	–	√
Euro. Barrier (*Info-reuse*)[1]	–	√	–	–
Amer. Vanilla. BT	√	–	–	–
Amer.2DMax (*Interp+S.Load*)	–	–	–	√

OpenCL + Workload Allocation

Kernels Runtime





	CPU	PHI	GPU	HET
Full Portf. BS	66.5	46.5	156.7	17.2
Full Portf. Heston	78.2	48.8	158.4	18.7

Summary

I. Overview: Portfolio Risk

II. OpenCL + Workload Allocation

III. Algorithmic Optimizations / Numerical Scheme

IV. Minimizing Device Global Memory

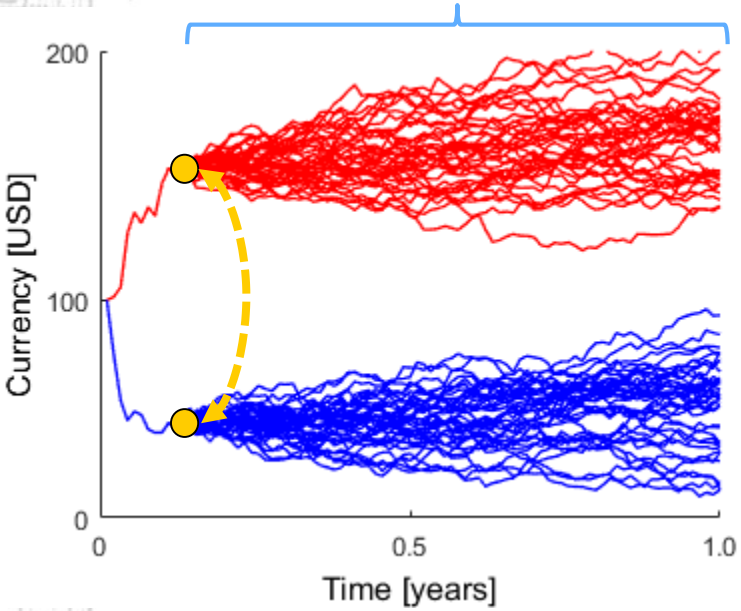
V. Conclusion

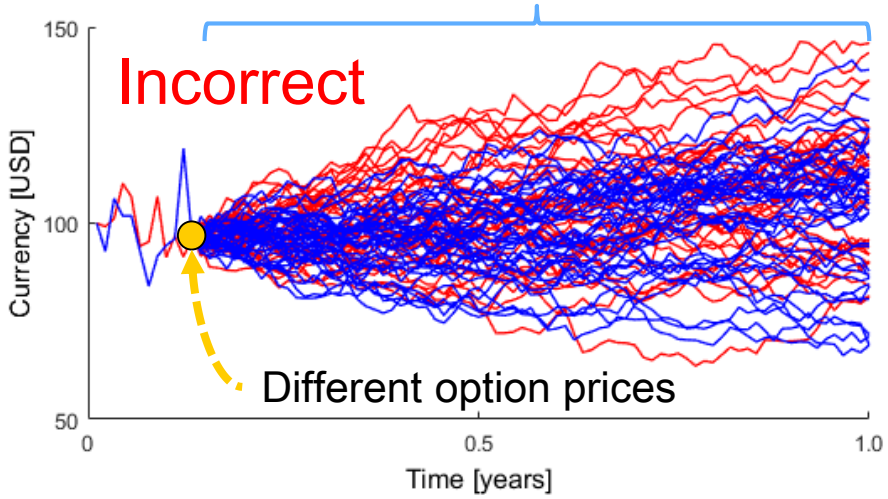
Algorithmic Optimizations

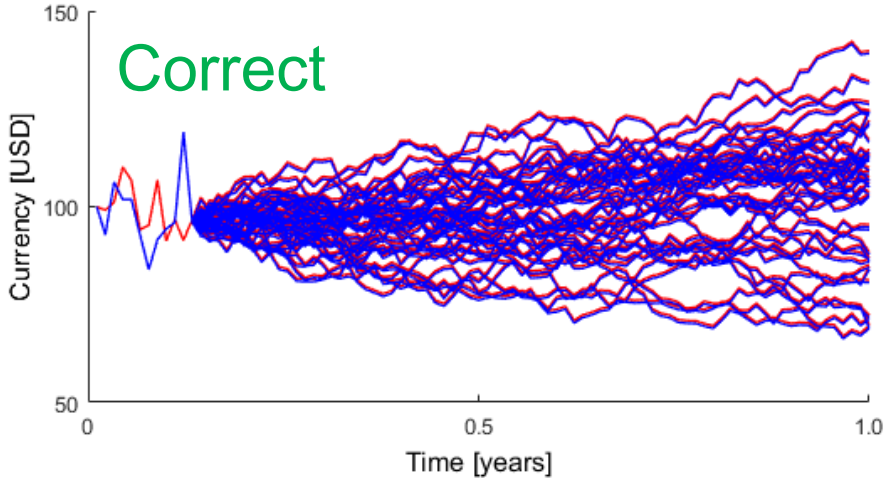
Random Numbers Reuse Approach

Nested MC-Based Risk Measurement of Complex Portfolios: Acceleration and Energy Efficiency
S. Desmettre, R. Korn, J. Varela, N. Wehn.
Risks Vol. 4, no. 4, pages 36, October, 2016.

(All parameters in the internal simulation remain constant)



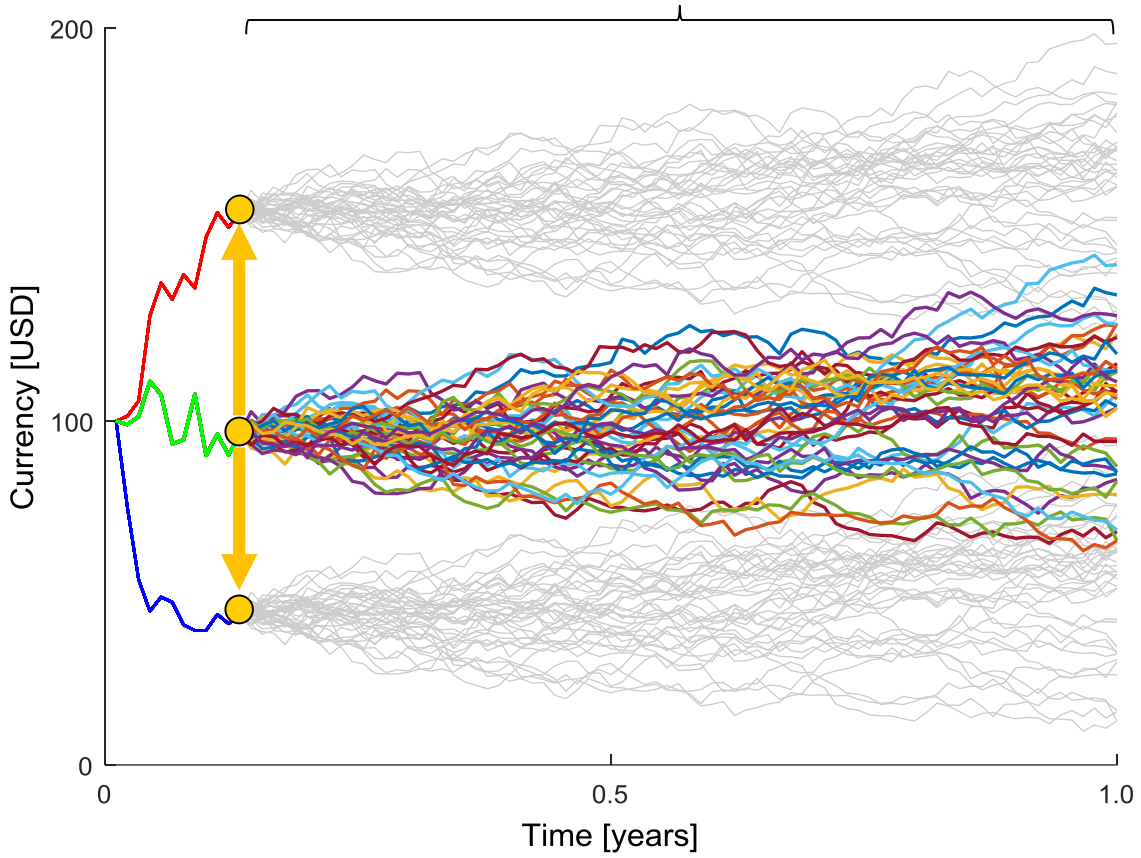




Algorithmic Optimizations

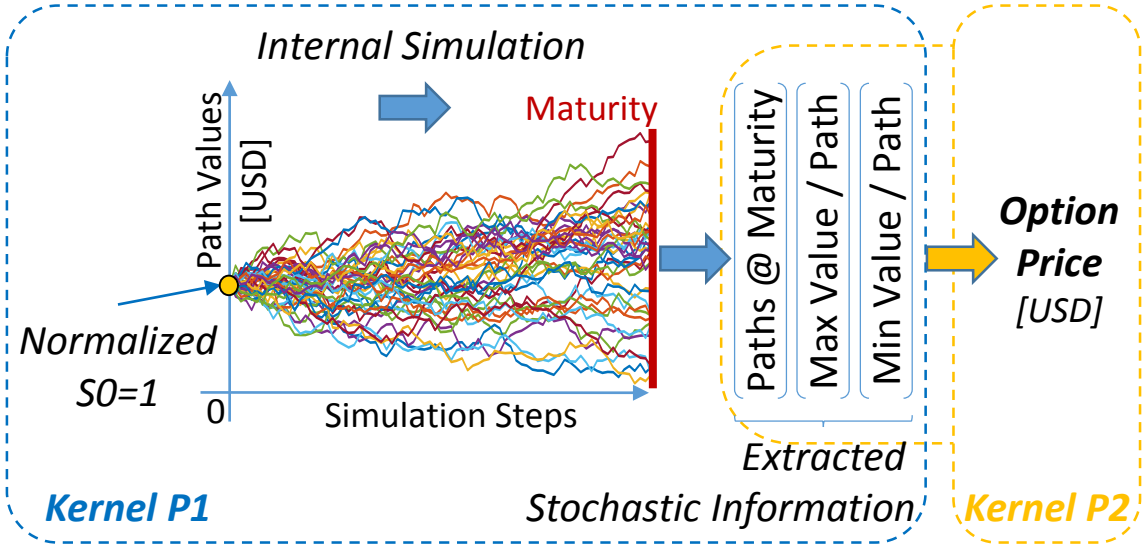
Paths-Reuse Approach (+Normalization)

(Note: All parameters in the internal simulation remain constant)



Algorithmic Optimizations

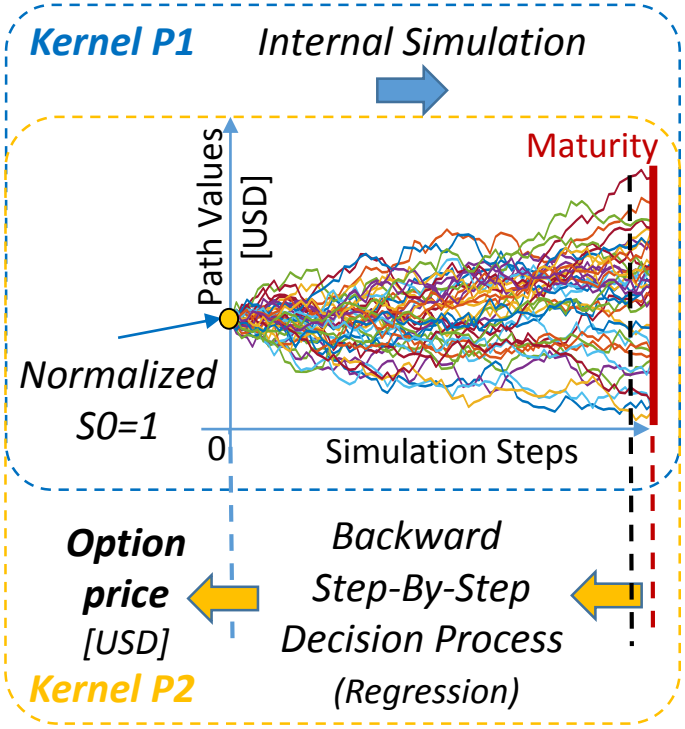
Paths-Reuse: European-style Options



➡ **Information-Reuse Approach**

Algorithmic Optimizations

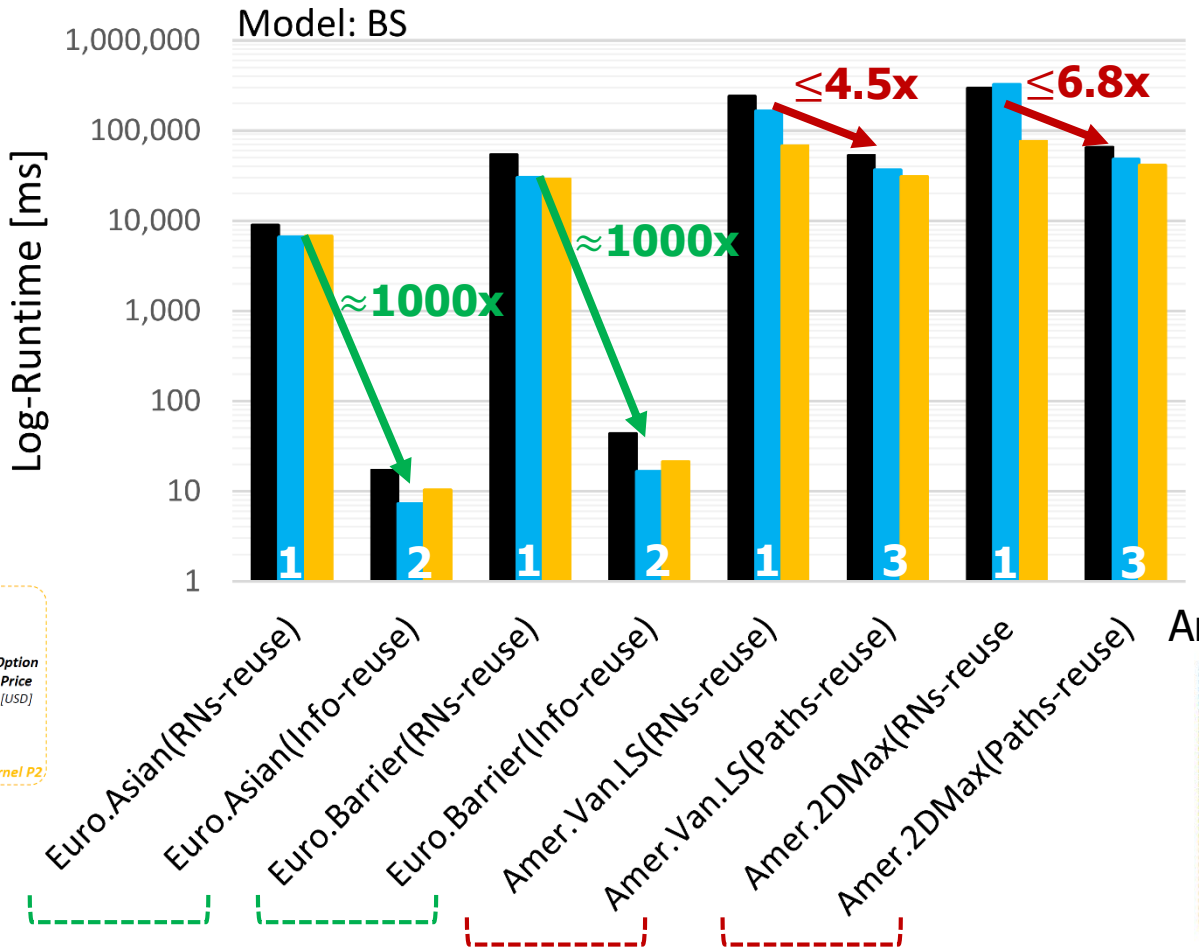
Paths-Reuse: American-style Options



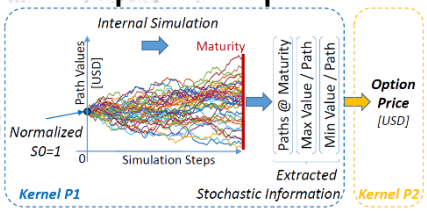
Algorithmic Optimizations

Runtime: Paths-/Info-Reuse (under BS model)

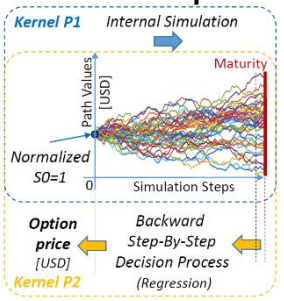
1. RNs-reuse
2. Info-reuse
3. Paths-reuse



European Options



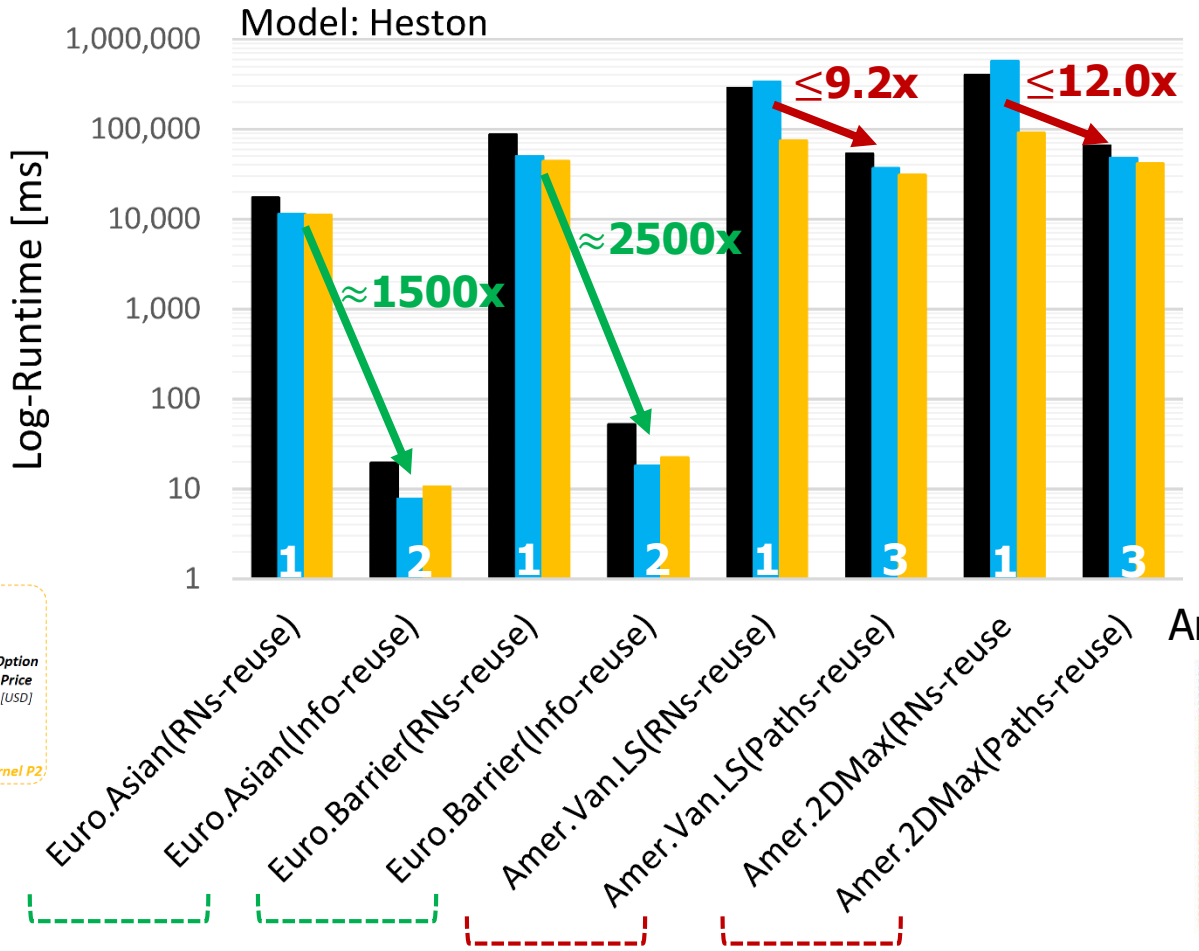
American Options



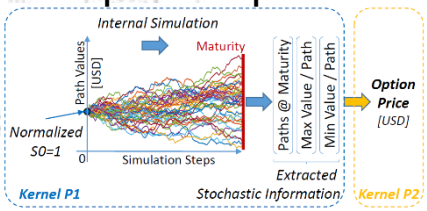
Algorithmic Optimizations

Runtime: Paths-/Info-Reuse (under BS model)

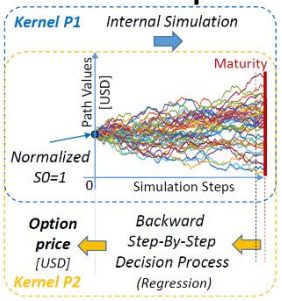
1. RNs-reuse
2. Info-reuse
3. Paths-reuse



European Options



American Options



Numerical Scheme

Interpolation:

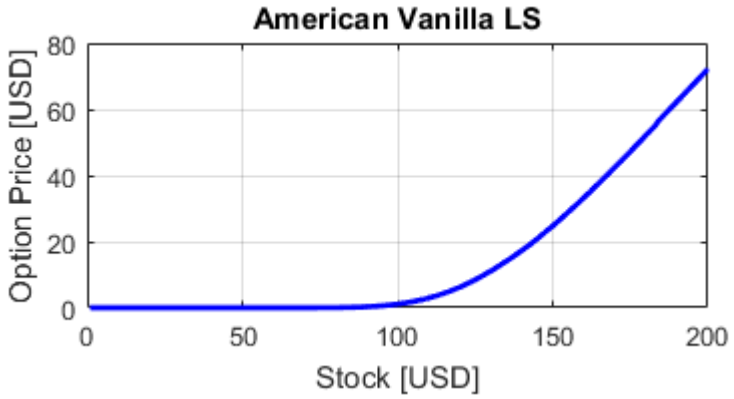
Fast, accurate and inelegant valuation of American options.
Adriaan Joubert and L.C.G. Rogers. 1997. In Proceedings of the Numerical Methods Workshop at the Isaac Newton Institute, April 1995, L.C.G. Rogers and D. Talay (Eds.). CambridgeS.

Kernel:

AmericanVanillaBT

AmericanVanillaLS

P1 → **P2**

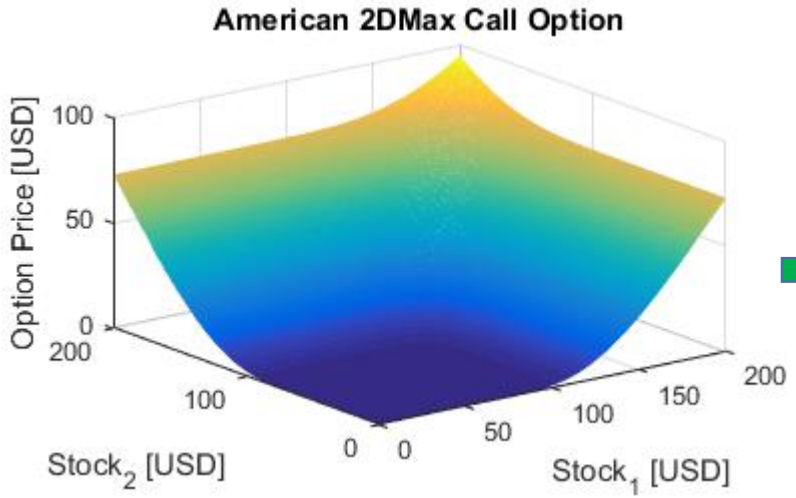


Kernel:
Interpolate 1D
Option Price
[USD]

Extension:

American2DMax

P1 → **P2**

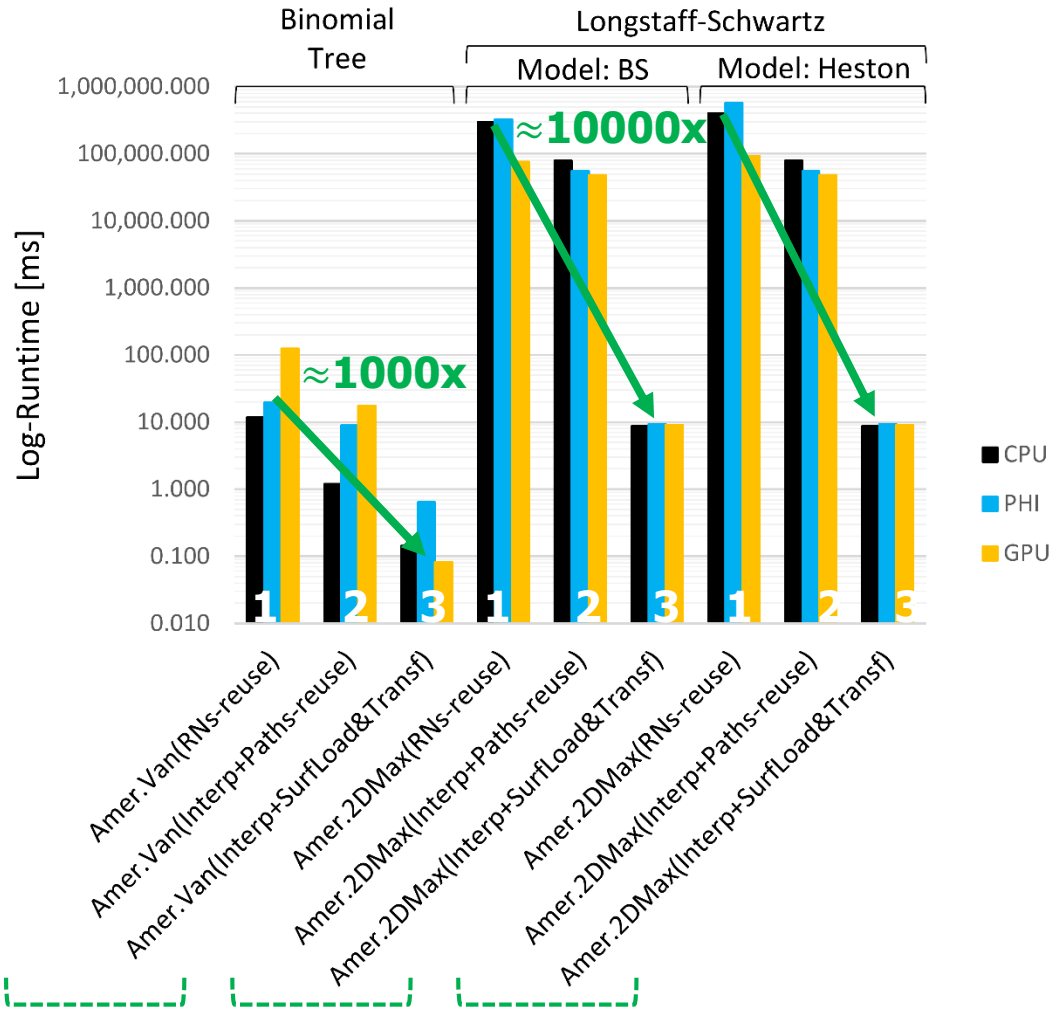


Interpolate 2D
Option Price
[USD]

Numerical Scheme

Runtime: Interpolation (American-style Options)

1. RNs-reuse
2. Interpolation+Paths-reuse
3. Interpolation+Surface Load&Transfer



Summary

I. Overview: Portfolio Risk

II. OpenCL + Workload Allocation

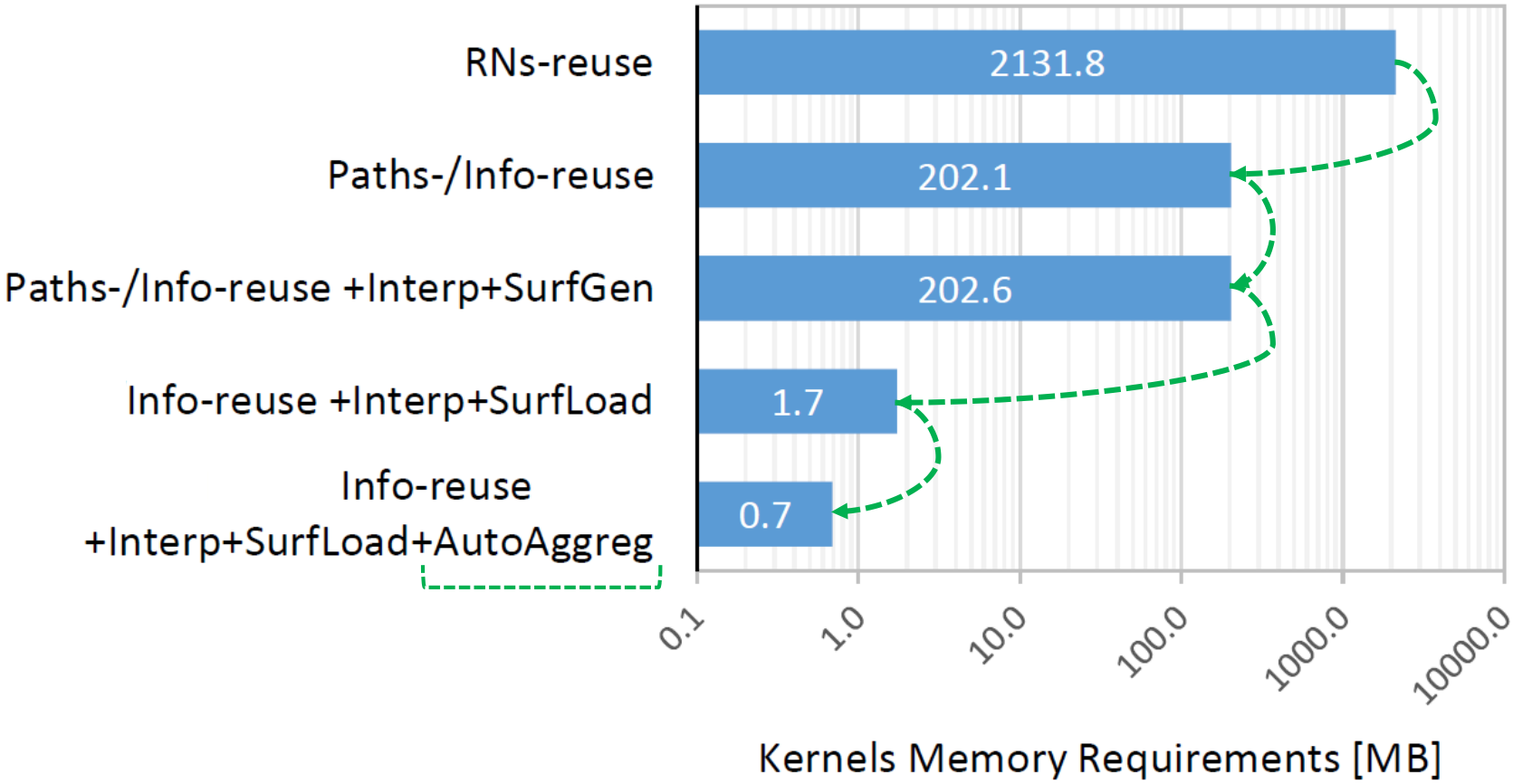
III. Algorithmic Optimizations / Numerical Scheme

IV. Minimizing Device Global Memory

V. Conclusion

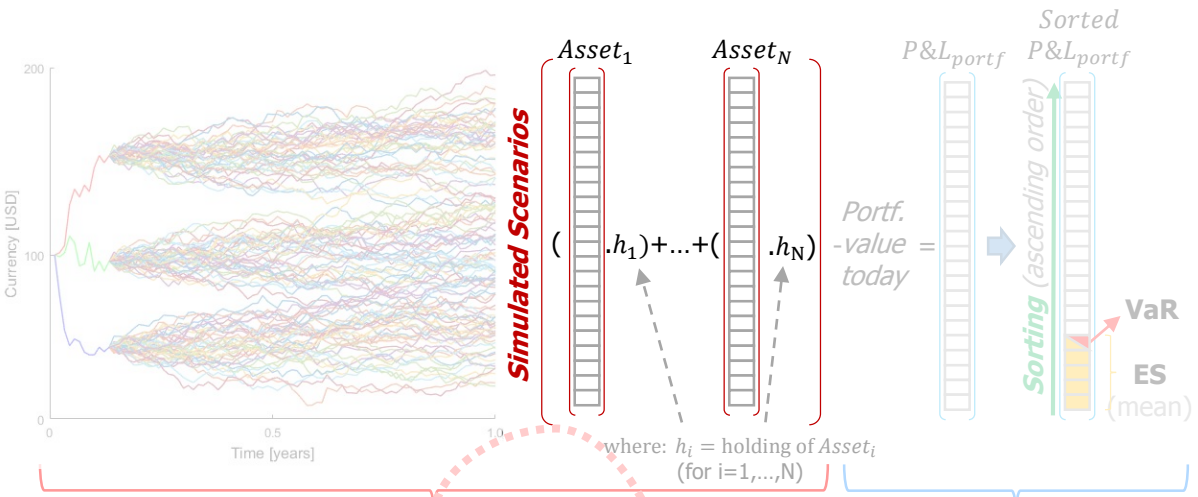
Minimizing Device Global Memory

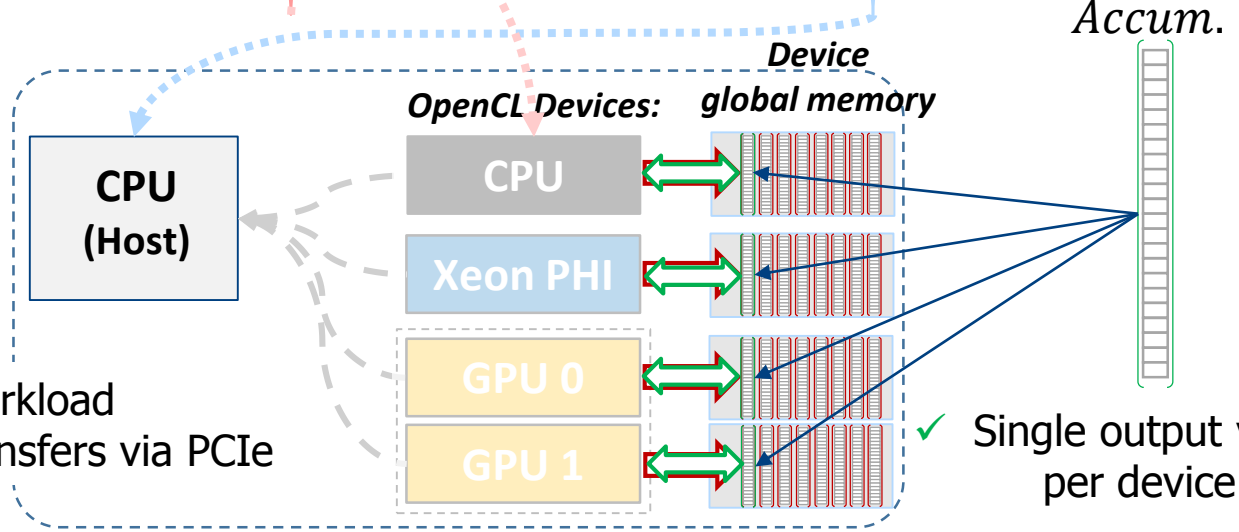
Device Global Memory Cost (Kernels)



Minimizing Device Global Memory

Autoaggregation





- ✓ Minimizes Host workload
- ✓ Minimizes data transfers via PCIe

✓ Single output vector per device

MICROELECTRONIC SYSTEMS DESIGN RESEARCH GROUP

Summary

I. Overview: Portfolio Risk

II. OpenCL + Workload Allocation

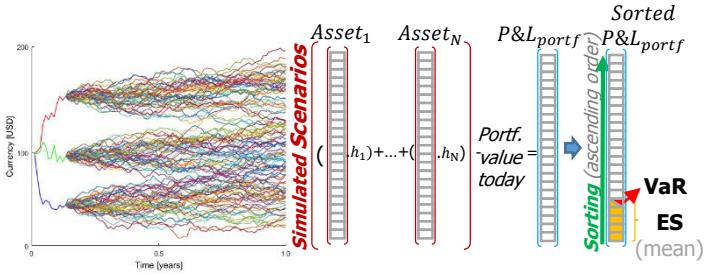
III. Algorithmic Optimizations / Numerical Scheme

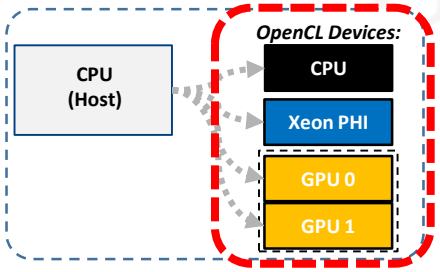
IV. Minimizing Device Global Memory

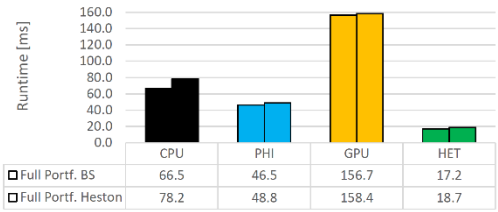
V. Conclusion

Conclusion

I. Overview: Portfolio Risk



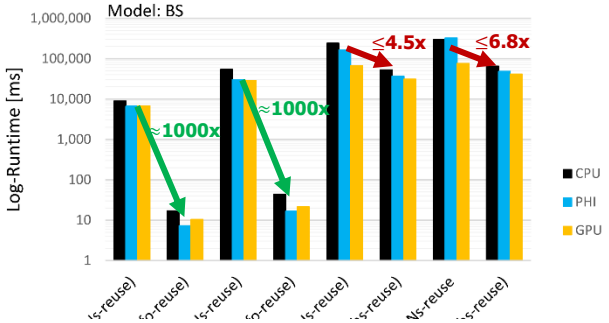


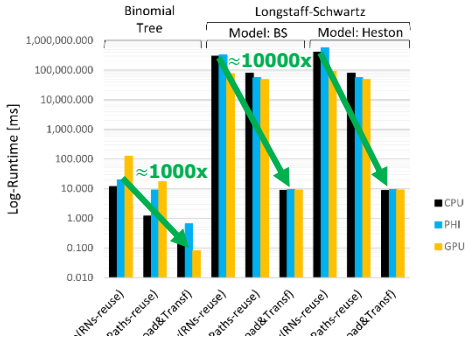


	CPU	PHI	GPU	HET
Full Portf. BS	66.5	46.5	156.7	17.2
Full Portf. Heston	78.2	48.8	158.4	18.7

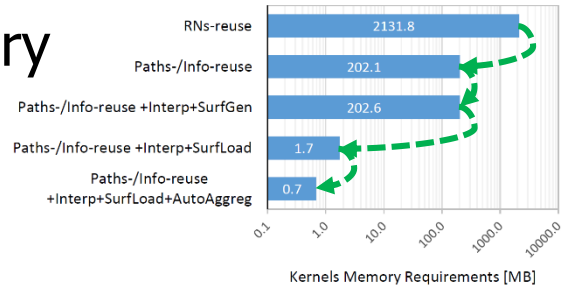
II. OpenCL + Workload Allocation

III. Algorithmic Optimizations / Numerical Scheme





IV. Minimizing Device Global Memory



- Portability (Numerical Results)

Questions ?

Thanks for your attention