Energy-Efficient Transprecision Techniques for Iterative Refinement

This work presents transprecision techniques for iterative refinement, which utilize various precision arithmetic dynamically according to numeric properties of the algorithm and computational latencies depending on precisions. The transprecision techniques were plugged into a mixed precision iterative refinement on an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0. The transprecision techniques brought further 2.0 - 3.4 X speedups and 3.0 - 4.1 X energy reductions to a mixed precision iterative refinement when double precision solution accuracy was required for forward error and a matrix size was ranged from 4K to 32K.

### Results

**Transprecision Techniques on**

an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0

**Test matrices:** Dense uniformly distributed random matrices

**With TTs,**

- TT 1 achieves an intermediate accuracy ($10^{1.0}$) quicker.
- TT 2 enables accuracy to leap from $10^{12}$ to $10^{16}$
- TT 3 removes the time cost for second dbl-dbl refinement

**Speedups with TTs**

- More iterations in total but, only 1 dbl-dbl iteration using TT 2 and TT 3

**Mixed-IR Runtime < Uni-IR**

- Mixed-IR Runtime $\propto O(n^{2.5})$
- Uni-IR Runtime $\propto O(n^{3.0})$

**Runtime with TTs < Mixed-IR**

- More Energy Saving!

**Less Energy with TTs**

- Less Energy Saving with TTs

**Transprecision Techniques brought further**

- 2.0 - 3.4 X Speedup
- 3.0 - 4.1 X Energy Reduction

---

**Abstract**

This project has received funding by the European Commission Horizon 2020 research and innovation programme under grant agreement No 732631 (OPRECOMP)

---

**Methodology**

**Algorithm Numeric Properties**

Transprecision Techniques

**Latencies depending on Precisions**

Plugging them into mixed precision iterative refinement

**Mixed-IR : Double precision accuracy for forward error**

**Approximation**

Step 1: LUPP

$LU \times x^{(1)} = \mathbf{b}$

($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Refinement**

Step 2: $A \times x^{(1)} = b$

($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Accuracy Check:**

$|\frac{||x^{(1)} - x^{(2)}||}{||x^{(1)}||} + \epsilon| < \epsilon \times ||\mathbf{b}||$

$O(n^3)$

$O(n^3)$

Step 3: $LU \times x^{(2)} = \mathbf{b}$

($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Step 4:** $x^{(i+1)} = x^{(i)} - (\epsilon) x^{(i) \times x^{(2)}}$

$O(n^3)$

$O(n^3)$

Go back to step 2

$O(n^3)$

$O(n^3)$

**Mixed-IR**

**Uni-IR**

**With TTs,**

- TT 1 achieves an intermediate accuracy ($10^{1.0}$) quicker.
- TT 2 enables accuracy to leap from $10^{12}$ to $10^{16}$
- TT 3 removes the time cost for second dbl-dbl refinement

**Speedups with TTs**

- More iterations in total but, only 1 dbl-dbl iteration using TT 2 and TT 3

**Mixed-IR Runtime $\propto O(n^{2.5})$**

- Uni-IR Runtime $\propto O(n^{3.0})$

**Runtime with TTs $<$ Mixed-IR**

- More Energy Saving!

**Less Energy with TTs**

- Less Energy Saving with TTs

**Transprecision Techniques brought further**

- 2.0 - 3.4 X Speedup
- 3.0 - 4.1 X Energy Reduction

**Speedups and Energy Savings**

### Impact of Precision

- **ALU Precision**
  - Lower
  - Higher

- **Shorter Wire**
  - Shorter ALUs
  - Higher Speedup

- **Higher Performance**
  - Higher Speedup
  - Higher Performance

---

**Transprecision Techniques**

- Plugging them into mixed precision iterative refinement

---

**Background**

Need some techniques for energy saving? Mixed Precision Method

**Parallel Computing with $m \times M$ Cores**

- $m \times m$ Speedup
  - $m \times m$ Power
  - $1 \times X$ Energy

**Mixed Precision Iterative Refinement without increasing cores**

- $n \times n$ Speedup
  - $1 \times X$ Power
  - $1/n \times X$ Energy

**Transprecision Techniques for Mixed precision Iterative Refinement**

- $(s \times n) \times X$ Speedup
  - $1 \times X$ Power
  - $1/(s \times n) \times X$ Energy

**Numerical Properties**

- **Numerical Properties (NP) and Transprecision Techniques (TT)**
  - **NP 1**
    - Start double for Step 2 and switch it to dbl-dbl when the convergence is saturated
  - **NP 2**
    - Inaccurate rounding errors in $r$ through Step 3
  - **NP 3**
    - Double precision accuracy guaranteed if $c_i = c_d$ and single precision accuracy for $x$ obtained using TT 2

---

**Acknowledgements**

JunKyu Lee, Hans Vandierendonck, Dimitrios S. Nikolopoulos
Centre for Data Science and Scalable Computing, Queen’s University of Belfast, UK

---

**Graphs and Figures**

- Accuracy and Runtime Trade-off
- Mixed-IR
- Uni-IR
- Trans-IR
- Trans-IR (Inner Loop)
- Mixed-IR
- Uni-IR
- Trans-IR

- Energy Consumption (AEMA LAPL)
- Mixed-IR
- Uni-IR
- Trans-IR

- Speedups and Energy Savings
- Mixed-IR
- Uni-IR
- Trans-IR

- Transprecision Techniques on
  - an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0
  - Test matrices: Dense uniformly distributed random matrices

- With TTs,
  - TT 1 achieves an intermediate accuracy ($10^{1.0}$) quicker.
  - TT 2 enables accuracy to leap from $10^{12}$ to $10^{16}$
  - TT 3 removes the time cost for second dbl-dbl refinement

- Speedups with TTs
  - More iterations in total but, only 1 dbl-dbl iteration using TT 2 and TT 3

- Mixed-IR Runtime $\propto O(n^{2.5})$
  - Uni-IR Runtime $\propto O(n^{3.0})$

- Runtime with TTs $<$ Mixed-IR
  - More Energy Saving!

- Less Energy with TTs
  - Less Energy Saving with TTs

- Transprecision Techniques brought further
  - 2.0 - 3.4 X Speedup
  - 3.0 - 4.1 X Energy Reduction
  - to Mixed-IR

---

**Impact of Precision**

- **ALU Precision**
  - Lower
  - Higher

- **Shorter Wire**
  - Shorter ALUs
  - Higher Speedup

- **Higher Performance**
  - Higher Speedup
  - Higher Performance

---

**Algorithm Numeric Properties**

Transprecision Techniques

**Latencies depending on Precisions**

Plugging them into mixed precision iterative refinement

---

**Mixed-IR : Double precision accuracy for forward error**

**Approximation**

Step 1: LUPP

$LU \times x^{(1)} = \mathbf{b}$

($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Refinement**

Step 2: $A \times x^{(1)} = b$ ($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Accuracy Check:**

$|\frac{||x^{(1)} - x^{(2)}||}{||x^{(1)}||} + \epsilon| < \epsilon \times ||\mathbf{b}||$

$O(n^3)$

$O(n^3)$

Step 3: $LU \times x^{(2)} = \mathbf{b}$ ($O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR)

$O(n^3)$, $c_d$ for Mixed-IR, $c_o$ for Uni-IR

**Step 4:** $x^{(i+1)} = x^{(i)} - (\epsilon) x^{(i) \times x^{(2)}}$

$O(n^3)$

$O(n^3)$

Go back to step 2

$O(n^3)$

$O(n^3)$