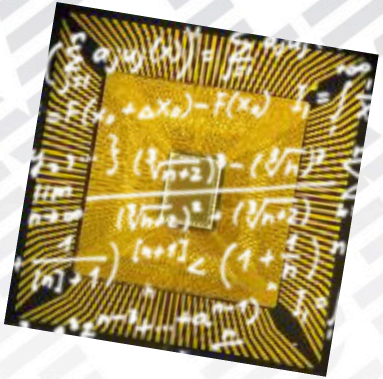
Decimal and Binary QP Precision Floating Point on IBM z13™



Cédric Lichtenau, Steven Carlough, Silvia Müller

IBM System Group - Processor Development

23rd IEEE Symposium on Computer Arithmetic

IBM

Outline

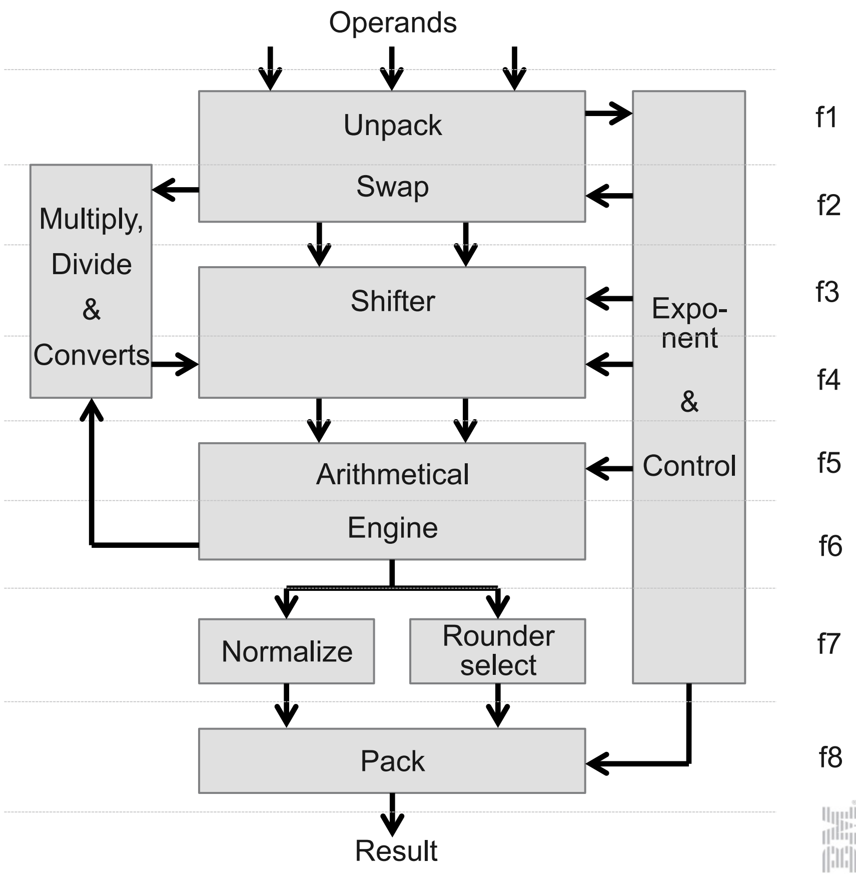
- Motivation
- Pipeline overview of the DQE
- Implementation highlights
- Performance results
- Conclusion

Workloads Requirement

- Decimal computation
  - Widely used for workload in Banking, Accounting and traditional Insurance code
  - Cobol, PL/1 and JAVA Big Decimal, C/C++
- Binary Quad Precision
  - Big Data Analytics, Risk Assessment code
  - Increased mathematical stability of the algorithms
  - Big ILOG installation: Quad precision in critical routines allows 18% faster convergence
- New hardware design point to support traditional and emerging workload
- Implementation
  - Area and power matters → Highly efficient design required
  - Combine and share logic between decimal and binary FP and BCD functions

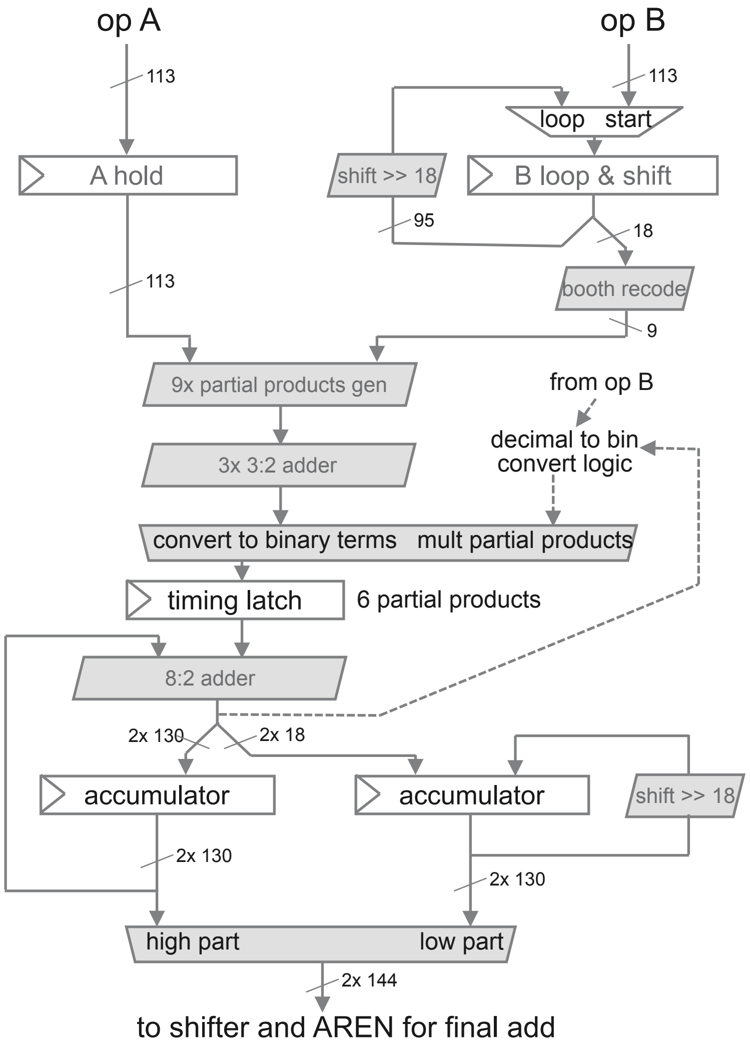
Decimal and Binary Quad Precision Engine

- 8 cycles pipeline for pipelined ops
- Full HW support for subnormal numbers
- Multi-cycles operations
  - Decimal Multiply and Divide
  - Binary Multiply
  - Converts Decimal<->Binary
- Supports 7 FP formats and BCD format
  - DFP: QP (34d), DP (16d), SP (7d)
  - BFP: QP (113b)
  - HFP: QP (28d), DP (14d), SP (7d)
  - BCD: 31d



DQE – Binary Multiplier

- Computes 18bits per cycle
- Booth encoding to reduce to 9 partial prods
- Accumulation in redundant format
- Final add, shift and round
- Circuit reused for converts from decimal to binary



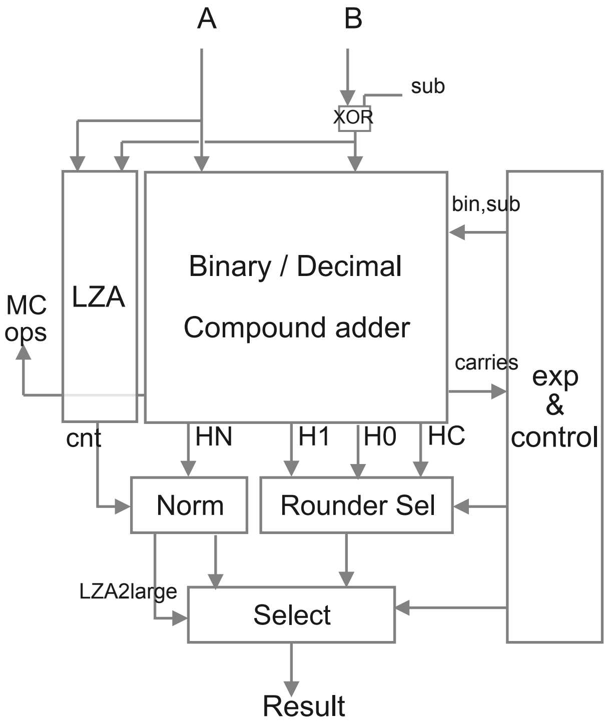
DQE – Arithmetical Engine

- QP compound adder (144bit wide)
- Supports binary, hex and decimal add/sub
  - Sign magnitude arithmetic via end-around-carry
  - Uses injection rounding

$$\mathrm{RRes} = \begin{cases} A+B & \text{if add} \\ A-B & \text{if sub and } (A > B) \\ B-A & \text{if sub and } (A <= B) \end{cases} = \begin{cases} H0/H1 & \text{if add} \\ H1/H0 & \text{if sub and (eac or bin)} \\ HC/H1 & \text{if sub and (!eac or !bin)} \end{cases}$$

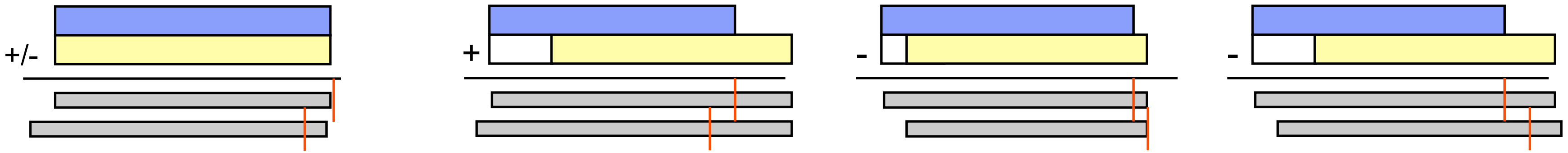
- Normalizer (bin only)

$$\mathrm{NRes} = \begin{cases} A+B & \text{if add or special} \\ A-B & \text{if sub and } (A > B) \\ B-A & \text{if sub and } (A <= B) \end{cases} = \begin{cases} H0 & \text{if add (deno) or special} \\ H1 & \text{if sub and bin and eac} \\ HC & \text{if sub and bin and !eac} \end{cases}$$



DQE – Rounder or Normalizer

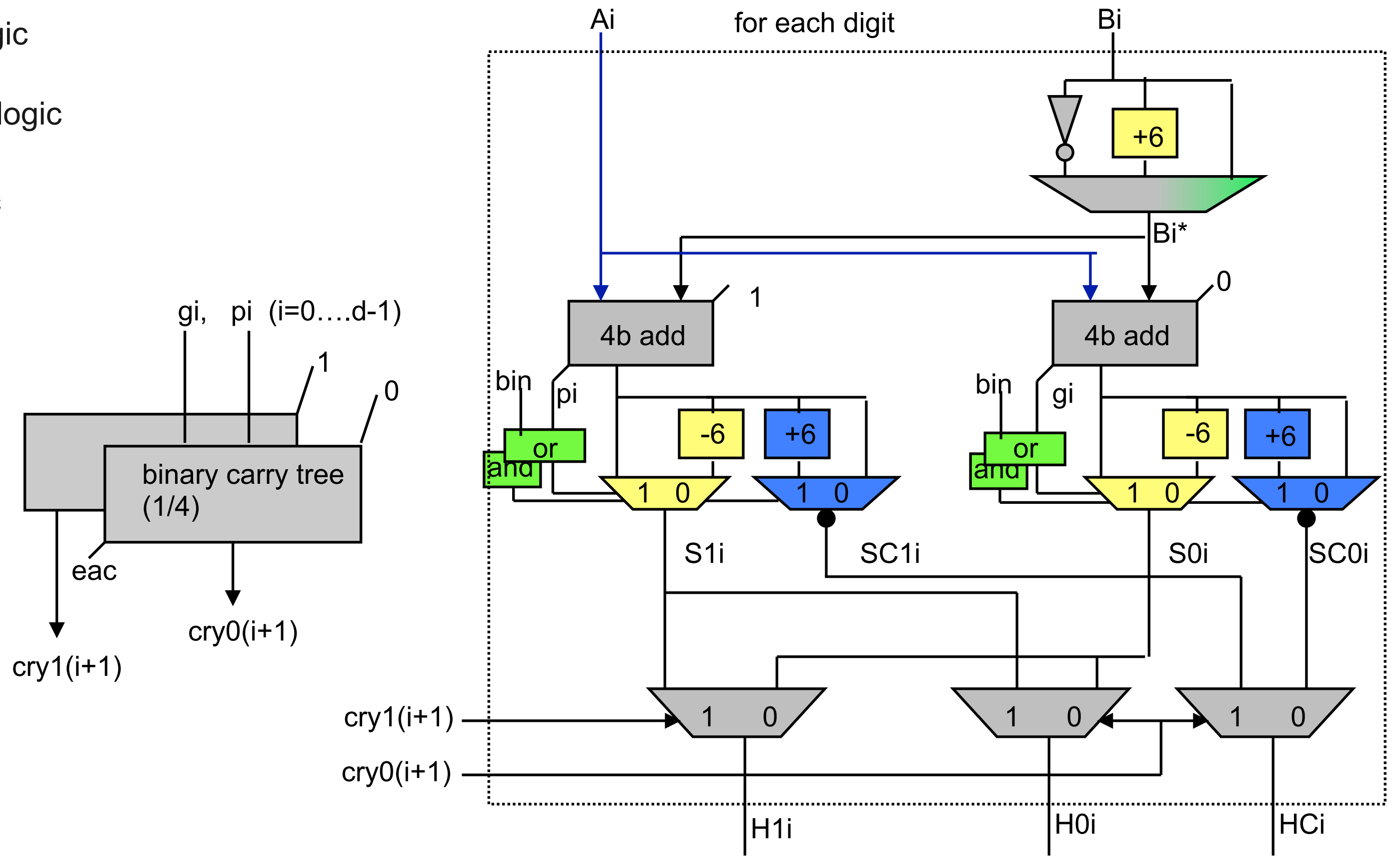
- Normalizer and rounder in parallel to save delay



<table>
<tr><th>eA = eB</th><th>eA = eB+1</th><th>eA > eB+1</th></tr>
<tr><td>Add: A, B ≥ NMIN
• Max 1 extra bit => RND
• A+B ≥ NMIN => no UNF</td><td colspan="2" rowspan="2">Add: A is normal, B get aligned
• Max 1 extra bit => RND
• A+B ≥ NMIN => no UNF</td></tr>
<tr><td>Add: A, B < NMIN
• Result max. 1.fff => exact, no OVF
• May UNF => NORM</td></tr>
<tr><td rowspan="2">Sub:
• No align, res ≤ A => exact, no OVF
• May UNF => NORM</td><td>Sub: Result keeps MSB
• B get aligned => RND
• A-B ≥ NMIN => no UNF</td><td rowspan="2">Sub:
• A is 1.fffff, B is 0.0fff => loose ≤ 1 bit
• Res max 1 bit shift => RND
• eA > eMIN+1 => no UNF</td></tr>
<tr><td>Sub: Result with cancelation
• B get aligned by 1 bit
• mantissa has max p bits => NORM</td></tr>
</table>

DQE – Sharing of the Adder Logic

- dec/bin shared logic
- dec only add/sub logic
- dec only sub logic
- bin only logic



DQE – Performance Comparison

- Binary FP QP performance results on the DQE compared to previous generation
  - Latency: # cycles between dependent instructions
  - CPI: # cycles before a new independent instruction can start

	Latency zEC12™	Latency z13™	CPI zEC12™	CPI z13™
Add/Sub	35	11	28	1.5
Multiply	55-97	23	48-90	7.5
Divide*	~165	49	~158	21
Sqrt*	~170	66	~163	24

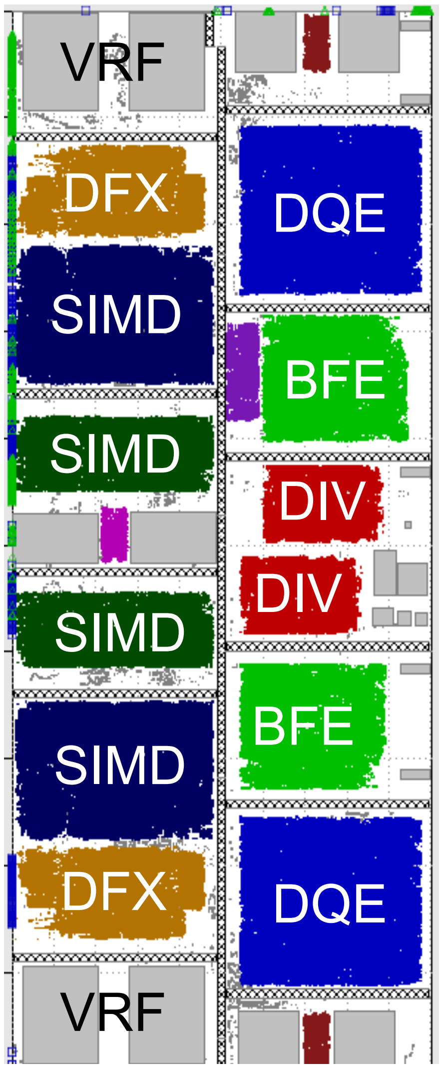
* Divide/Square Root executed in the Divide Engine, not in the DQE.

- Separated Div/SQRT and DQE engine to allow parallel execution
- About 3x better latency and 7-20x better CPI over zEC12™

Summary

- Highly efficient Decimal and Binary Quad Precision Design
- Shared logic supporting 8 different data types
- Runs at 5GHz in 22nm SOI Technology
- Widely improved performance over previous generation
- Protected by residue or parity to achieve high reliability

- Total area of the Vector FP Unit: 3.9 mm^2



Special Notice

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM, the IBM logo, ibm.com AIX, AIX (logo), AIX 5L, AIX 6 (logo), AS/400, BladeCenter, Blue Gene, ClusterProven, DB2, ESCON, i5/OS, i5/OS (logo), IBM Business Partner (logo), IntelliStation, LoadLeveler, Lotus, Lotus Notes, Notes, Operating System/400, OS/400, PartnerLink, PartnerWorld, PowerPC, pSeries, Rational, RISC System/6000, RS/6000, THINK, Tivoli, Tivoli (logo), Tivoli Management Environment, WebSphere, xSeries, z/OS, zSeries, Active Memory, Balanced Warehouse, CacheFlow, Cool Blue, IBM Watson, IBM Systems Director VMControl, pureScale, TurboCore, Chiphopper, Cloudscape, DB2 Universal Database, DS4000, DS6000, DS8000, EnergyScale, Enterprise Workload Manager, General Parallel File System, , GPFS, HACMP, HASM, IBM Systems Director Active Energy Manager, iSeries, Micro-Partitioning, POWER, PowerLinux, PowerExecutive, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Everywhere, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), Power Systems Software, Power Systems Software (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, POWER7+, Systems, System i, System p, System p5, System Storage, System z, TME 10, Workload Partitions Manager and X-Architecture are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries.

A full list of U.S. trademarks owned by IBM may be found at: http://www.**ibm.com**/legal/copytrade.shtml.

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Thank you

Questions?