Hewlett Packard Enterprise

Georgia Tech

Oak Ridge National Laboratory

Habanero Labs

Los Alamos National Laboratory

Heading Towards Energy Savings or Just Performance: Case-Study on Message Aggregation-Based Runtimes

Shubhendra Pal Singhal, Aaron Welch, **Nathan Wichmann**, Akihiro Hayashi, Oscar Hernandez, Bradford L. Chamberlain, Steve Poole, Vivek Sarkar

Introduction

Goal: Build a distributed application, which

Sends large volumes (~billions) of short messages (8-32 bytes) to remote processes using **aggregation**.

Is irregular and sparse.

State-of-the-art frameworks *Chapel* and *Conveyors* solve this with:

“Gain performance while maintaining productivity”

Challenge: But what about the **energy bills** incurred?

Energy is a serious problem

An official website of the United States government Here's how you know

U.S. DEPARTMENT of ENERGY

About · Our Mission · Topics · Funding Opportunities · Consumer Savings · Services & Opportunities

Energy.gov › DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers

DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers

Domestic Energy Usage from Data Centers Expected to Double or Triple by 2028, DOE Continues to Accelerate Development and Deployment of Solutions to Meet Growing Demand

Energy.gov

December 20, 2024

3 min

THE WHITE HOUSE
WASHINGTON

NEWS · GALLERY · LIVESTREAM · INVESTMENTS · SAVE AMERICA · WH WIRE · CONTACT

FACT SHEETS

Fact Sheet: President Donald J. Trump Advances Energy Affordability with the Ratepayer Protection Pledge

The White House | March 4, 2026

ENSURING TECH COMPANIES BUILD, BRING, OR BUY POWER FOR DATA CENTERS: Today, President Donald J. Trump brought the leading AI companies and hyperscalers together to sign the Ratepayer Protection Pledge, ensuring they protect Americans from electricity price hikes due to data center energy requirements now and in the long run, take action to further strengthen the grid, and ensure that all Americans benefit from the oncoming technological boom.

ENERGY IN AMERICA Published July 17, 2025 8:20am EDT | *Updated July 17, 2025 10:09am EDT*

Electricity prices may spike due to AI demand if US doesn't boost energy output, White House warns

White House sounds urgent alarm on looming energy crisis as AI technology growth accelerates

By Danielle Wallace, Edward Lawrence | FOXBusiness |

Add Fox Business on Google

U.S. Stock Market Quotes
Quotes displayed in real-time

THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Audio | Latest Headlines | Puzzles | More

World · Business · U.S. · Politics · Economy · Tech · Markets & Finance · Opinion · Free Expression · Arts · Lifestyle · Real Estat

BUSINESS • ENERGY & OIL Follow

U.S. Needs More Power for AI—but Critical Equipment Is Pricey and Scarce

Tariffs create new challenges for an industry overwhelmed by the sheer number of projects in line to connect to the nation's power grid

By Jennifer Hiller Follow and Katherine Blunt Follow

April 24, 2025 at 8:55 am ET

Goals of the paper

- ❖ Does aggregation yield energy-savings or is it an energy-hungry technique?
- ❖ Whether this deduction differs across frameworks, and if so, why?
- ❖ How the energy behavior of aggregation evolves as the scale of the system and problem size increase?

Methodology

Software: Weak Scaling for Histogram, Index Gather and Radix Sort for,

1. Chapel (un-aggregated version)
2. Chapel (aggregated version)
3. Conveyors (un-aggregated version)
4. Conveyors (aggregated version)

Datasets: Generated uniformly at random of size [64M…6400M] per node.

Hardware: *Frontier* at Oak Ridge National Laboratory. Each Frontier compute node consists of [1x] 64-core **AMD** "Optimized 3rd Gen **EPYC CPU**" (with 2 hardware threads per physical core) with access to 512 GB of DDR4 memory. The nodes are connected with [4x] **HPE Slingshot-11** 200 Gbps (25 GB/s) NICs, providing a node-injection bandwidth of 800 Gbps (100 GB/s).

Methodology: Measure the total time, energy consumed by Omnistat, and Derive GPS-UP (greenup, powerup, speedup) graphs.

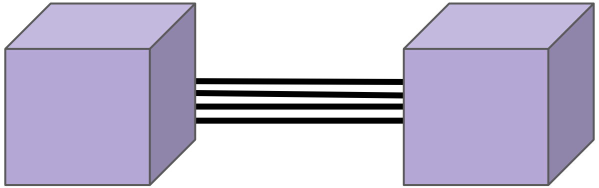
Overview of runtimes

Component	Conveyors	Chapel
Execution Model	single-threaded multi-process	multi-threaded multi-process
Buffer Allocation	per-PE (core) basis	per-locale (NIC) basis

Table 1: Differences between Conveyors and Chapel runtime.

Overview of Energy Measurement

Energy = Power * Time

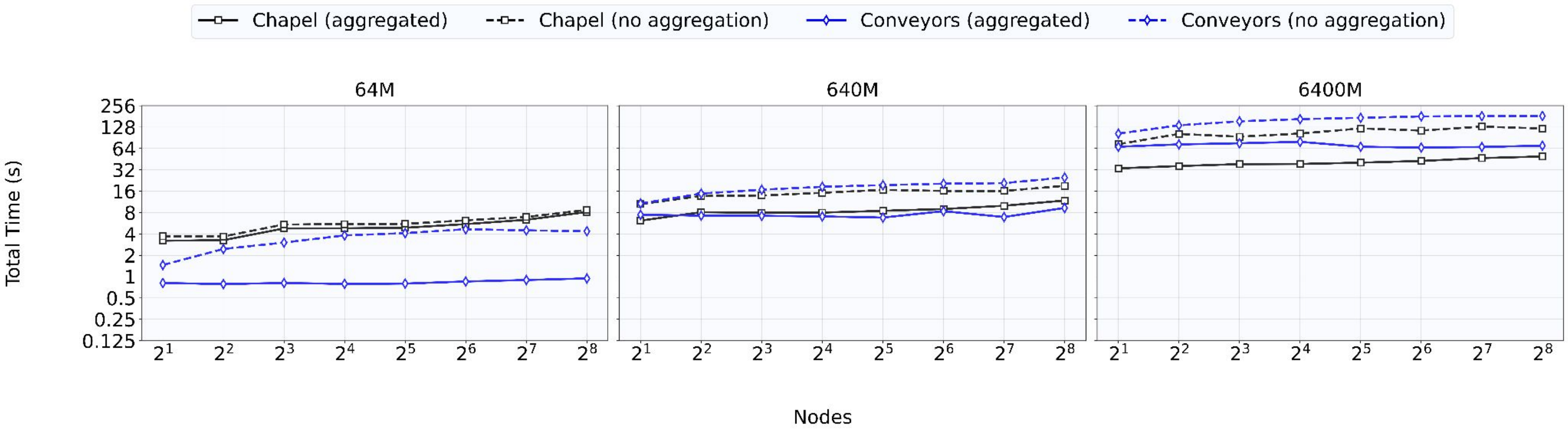


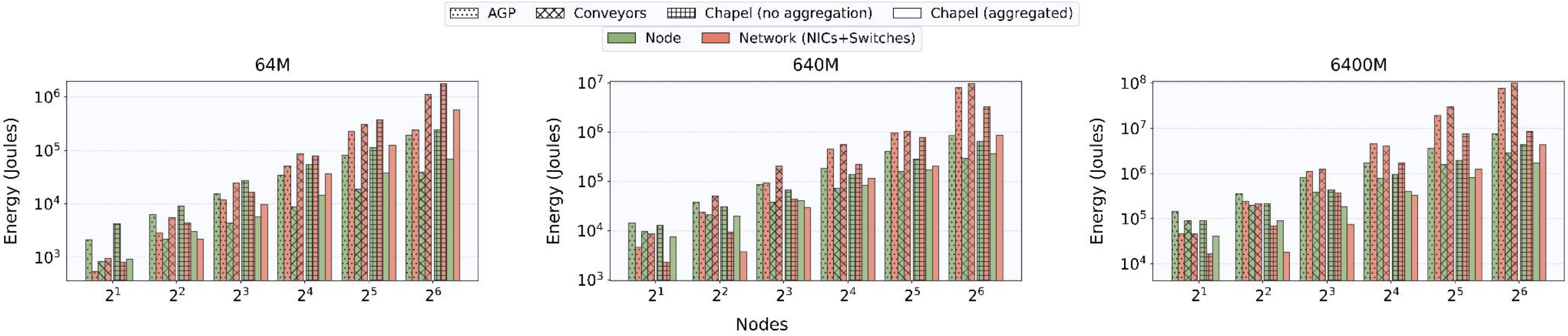
Energy of Node is measured via cray_pm:::PM_ENERGY:NODE, which includes CPU, memory & un-core parts.

Energy spent on wire is measured via **AMD Omnistat tool**.

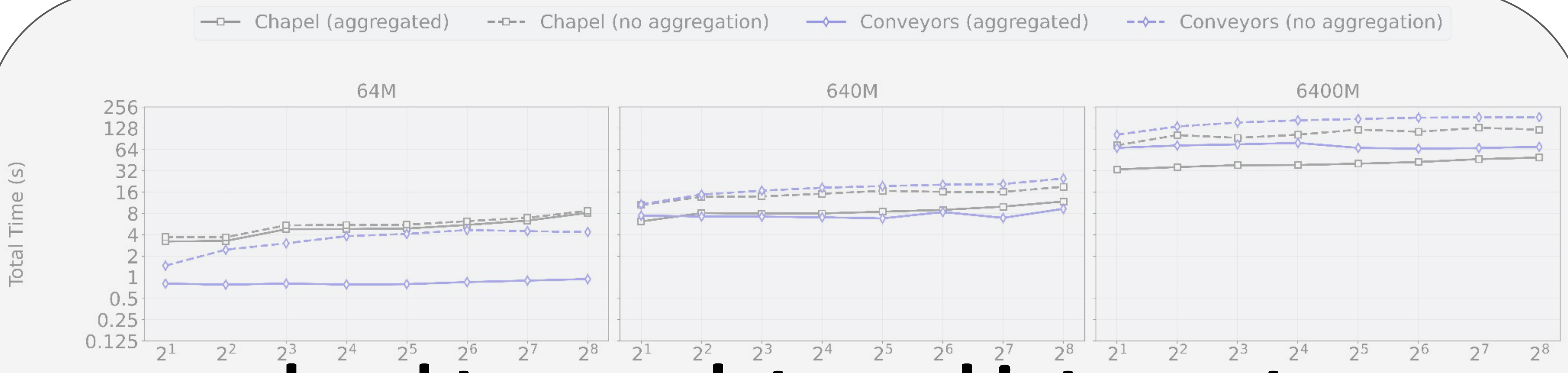
- Power draw by all four NICs on a node is 60W.
- Time spent on wire is calculated as `(△tx_ok_octets + △rx_ok_octets)/effective-bandwidth` where `delta(△)` is measured from start till the end. We use 10ms as query interval time.

Weak Scaling: Radix Sort

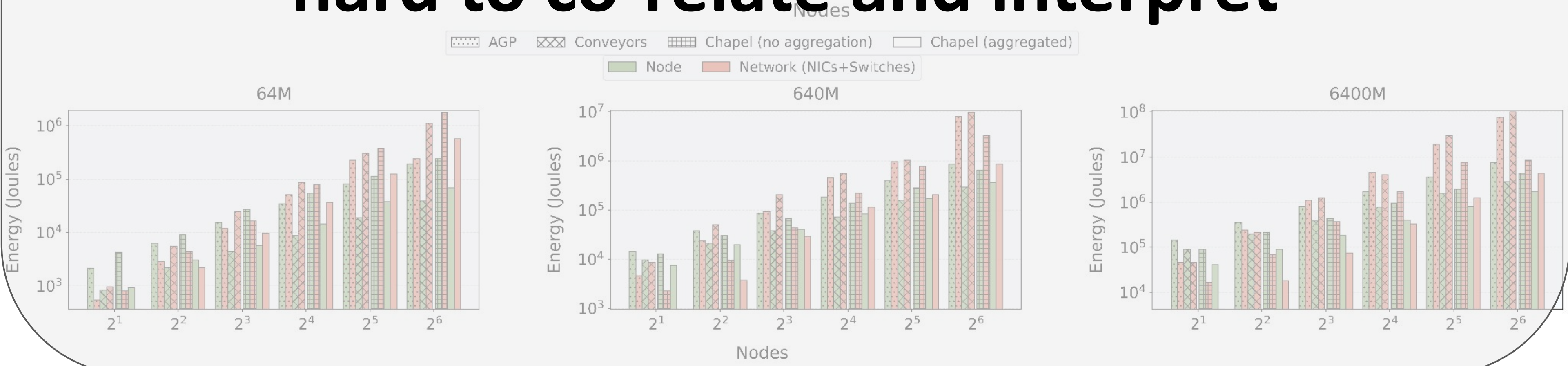




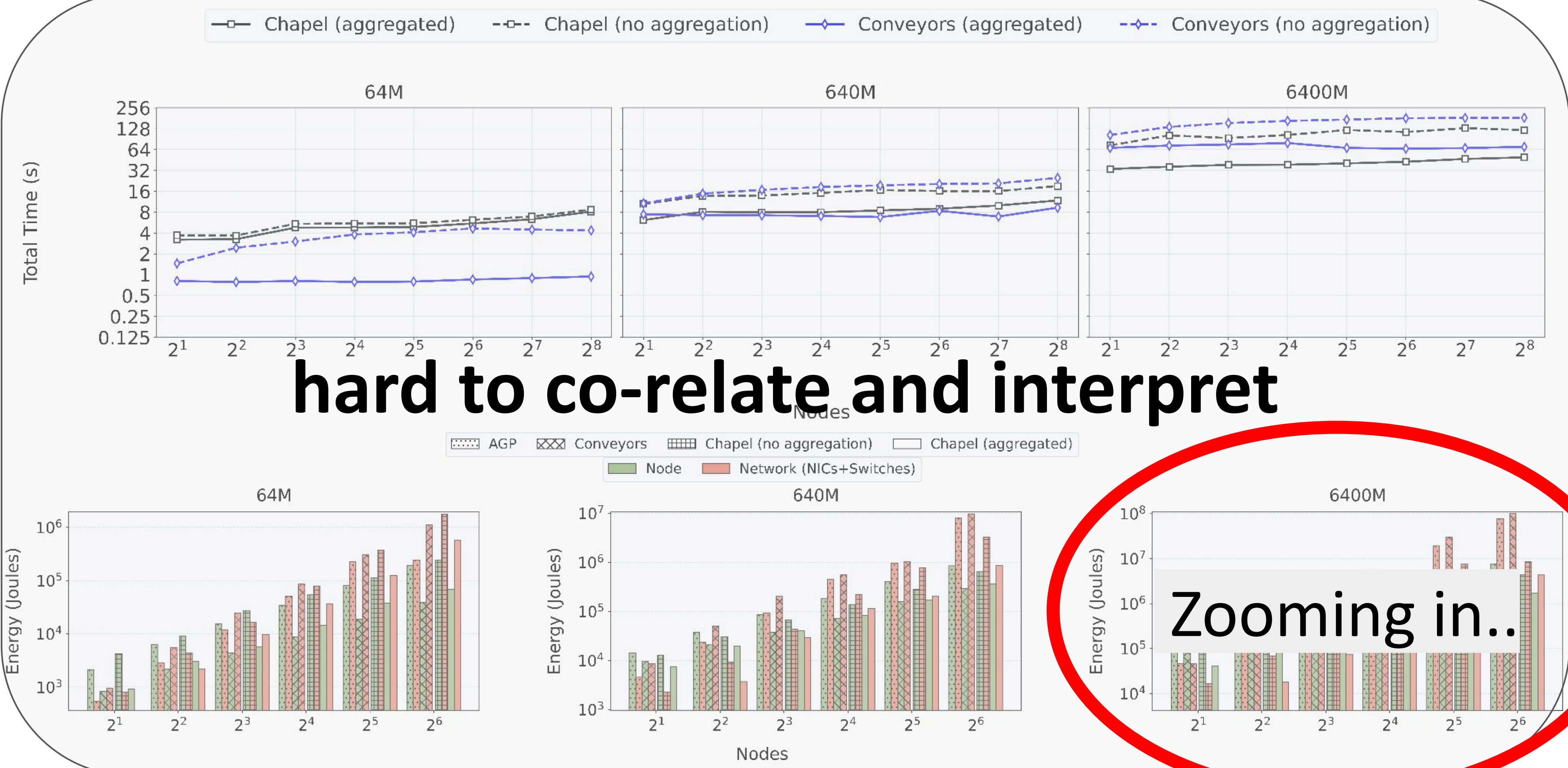
Weak Scaling: Radix Sort



hard to co-relate and interpret



Weak Scaling: Radix Sort

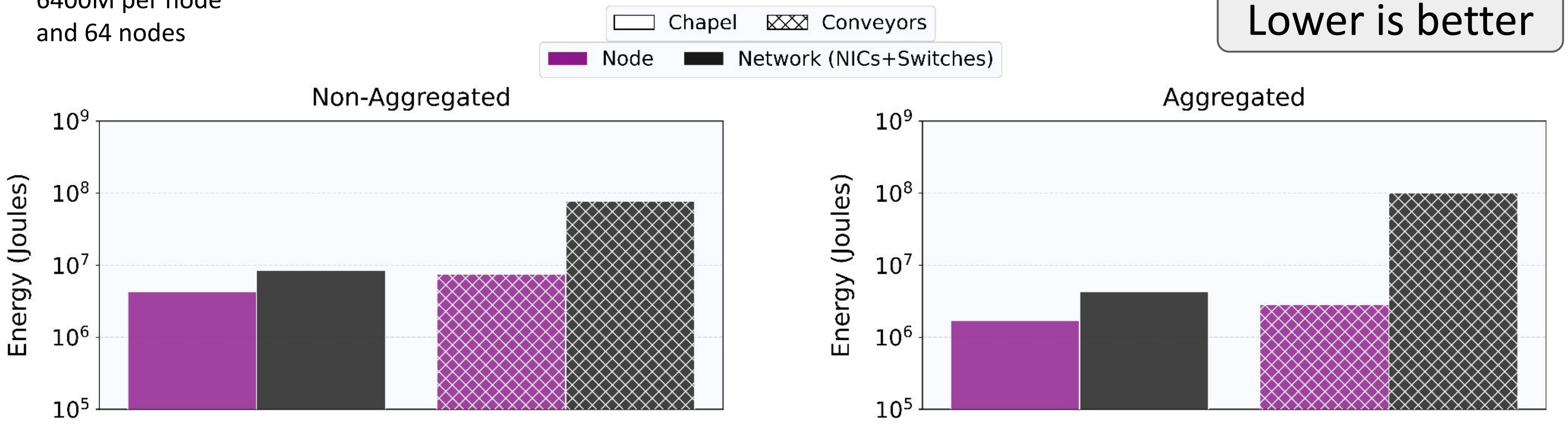


Weak Scaling: Radix Sort

End-to-End Energy Consumption

6400M per node
and 64 nodes

Lower is better



Non-Aggregated Version

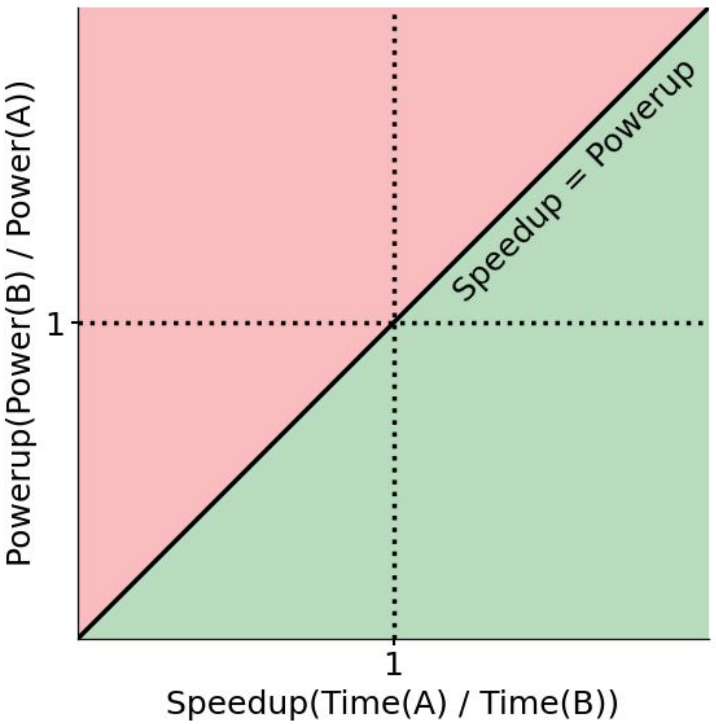
Energy(Chapel) ≤ Energy(Conveyors)

Aggregated Version

Energy(Chapel) < Energy(Conveyors)

Energy (NICs and switches) > Energy (Node)

Overview of GPS-UP plots



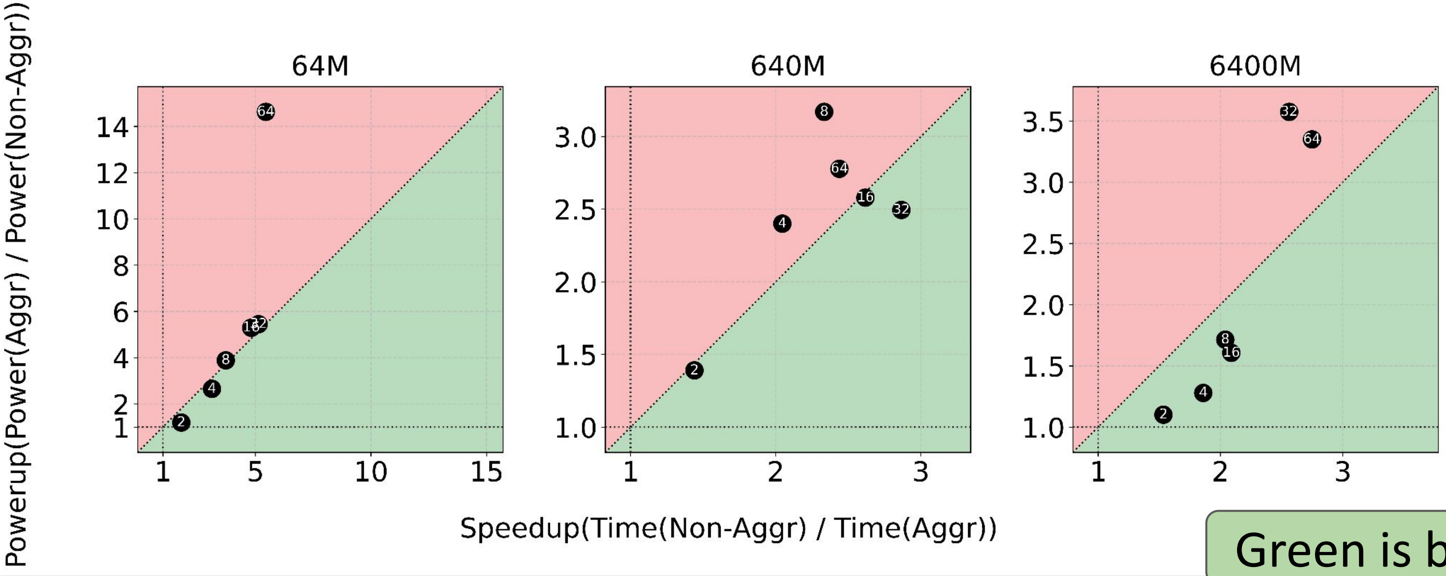
SpeedUp = $Time_A$: $Time_B$ *(measured empirically)*
GreenUp = $Energy_A$: $Energy_B$ *(measured empirically)*
PowerUp = Speedup : GreenUp *(derived)*

Green Area represents **energy savings.**
Red Area represents **energy-hungry**.

Why GPS-UP is a better metric than Energy Delay Product (EDP), refer to S. Abdulsalam, Z. Zong, Q. Gu and Meikang Qiu, "Using the Greenup, Powerup, and Speedup metrics to evaluate software energy efficiency," 2015 Sixth International Green and Sustainable Computing Conference (IGSC), Las Vegas, NV, USA, 2015, pp. 1-8, doi: 10.1109/IGCC.2015.7393699. keywords: {Software;Measurement;Optimization;Energy consumption;Runtime;Hardware;Software algorithms;software energy efficiency;software power measurement;software optimization;software evaluation metrics},

GPS-UP: Radix Sort

Conveyors (Non-aggregated vs Aggregated)

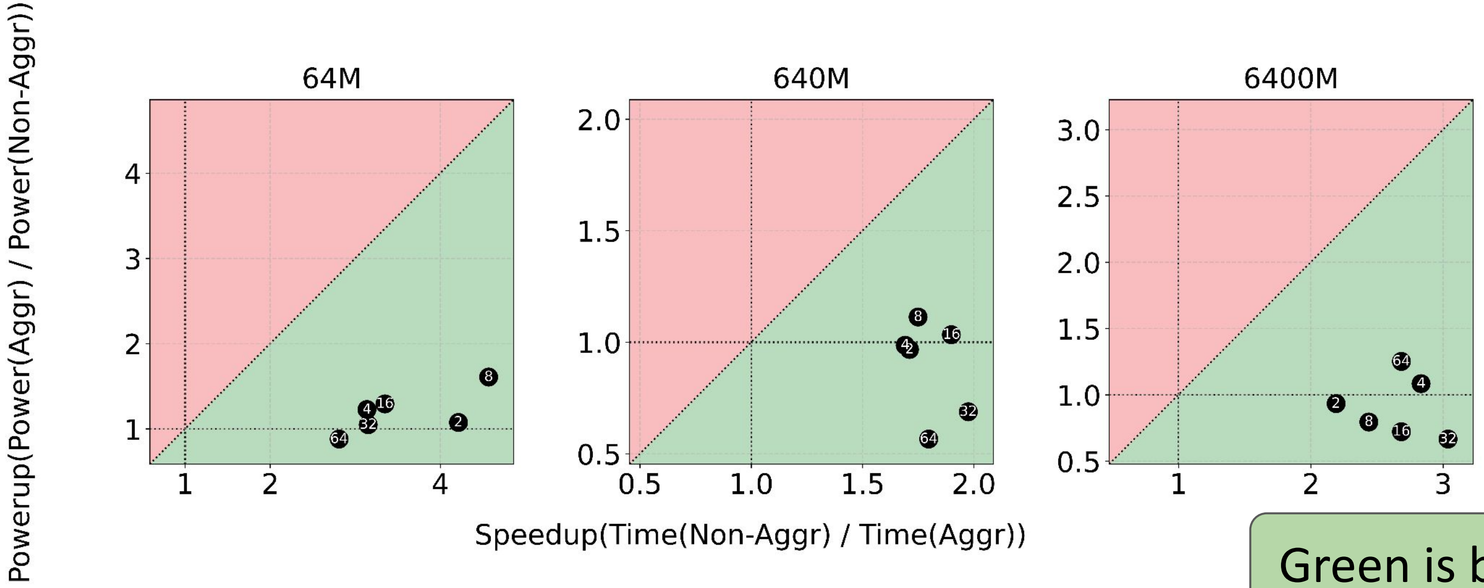


Green is better

As data scales increase, aggregation in Conveyors moves from energy-hungry regime to saving energy regime, and consumes more power.

GPS-UP: Radix Sort

Chapel (Non-aggregated vs Aggregated)

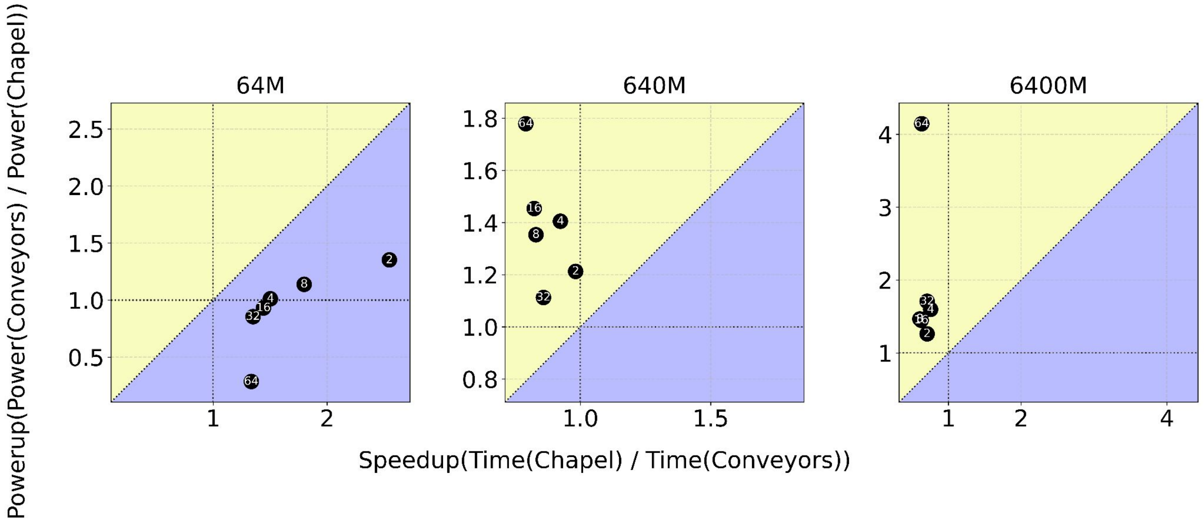


Green is better

Aggregation in Chapel saves energy but consumes more power for low data scale (64M). For the rest, aggregation saves energy!

GPS-UP: Radix Sort

Conveyors and Chapel (Non-aggregated)

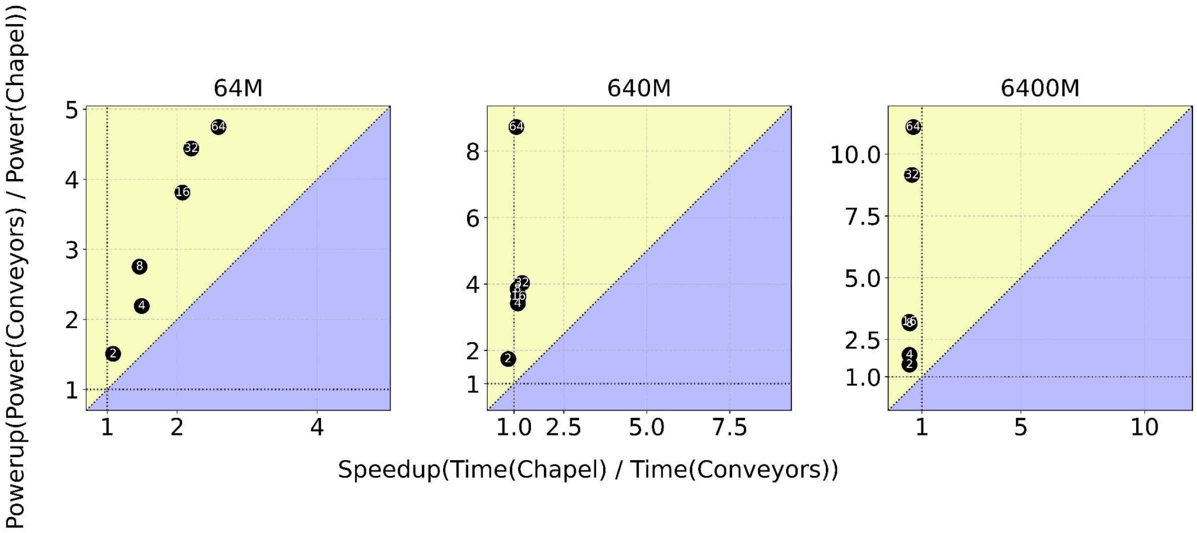


Yellow is better for Chapel

Blue is better for Conveyors

Non-aggregated: Conveyors saves energy for low data scale (64M).
For the rest, Chapel is energy-efficient .

GPS-UP: Radix Sort
Conveyors and Chapel (Aggregated)



Yellow is better for Chapel

Blue is better for Conveyors

Aggregated: Chapel is energy-efficient.

We tested more applications: Histogram and Index Gather and obtained similar insights on energy efficiency.

Note that the intent of this paper is not to recommend any one framework, but to focus on which developmental principle of programming saves energy!

Inference

Aggregation always ***saves time***.

Aggregation in Chapel always ***saves energy and time***.

→ *Conveyors is a single-threaded multi-process model* which mandates the need for **multiple sorting and communication stages** to combat the problem of memory insufficiency for aggregation per core.

→ *Chapel is a multi-threaded multi-process model* which incurs only a **single sorting and communication stage** since the model does not face memory problems, **and optimizes communication using remote caches.**

Since network is dominant consumer of energy, Chapel ***saves energy*** *relative to Conveyors.*

Conclusion

*“runtimes **must adopt aggregation** along-with **one-hop routing** using **remote cache** and **multi-threaded multi-process execution model** to yield energy savings, given that the applications are irregular, sparse, and network-bound, and the cluster is equipped with Slingshot-11 topology”*

We would like to extend our gratitude to **Wael Elwasif from Oak Ridge National Laboratory**, who helped us set up the environment for Chapel on Frontier; **Karl W. Schulz, Jorda Polo, and Bruno Villasenor Alvarez from AMD**, who assisted us with the workings of Omnistat on Frontier; **Michael Ferguson from HPE**, who informed us about optimized versions of the radix sort implementation in Chapel; and **Engin Kayraklioglu from HPE**, who guided us through Chapel's optimizations for Frontier. Lastly, we would like to acknowledge the **seminars conducted by the Chapel community at HPE**, which helped us brainstorm our initial findings on execution time, power, and energy efficiency.

Thank you
Q/A time!