Tom Scogland- Virginia Tech & Green500,
Erich Strohmaier- LBNL & TOP500,
Michael Patterson- Intel and The Green Grid,
Torsten Wilde- LRZ,
Aaron Anderson- NCAR,
Jean-Philippe Nomine, CEA,
Frederick Lefebvre, Calcul Quebec/University Laval,
Buddy Bland, ORNL with
Natalie Bates, EE HPC Working Group
And others on the Compute System Metrics Team

SETTING TRENDS FOR ENERGY-EFFICIENT SUPERCOMPUTING

ISC13 BoF; June 2013; Leipzig, Germany

Agenda

- Introduction – Tom Scogland 10min 2:15- 2:25
- LRZ – Torsten Wilde 8min 2:26- 2:34
- NCAR – Aaron Andersen 8min 2:35- 2:43
- CEA – Jean-Philippe Nomine 8min 2:44- 2:52
- Calcul Quebec – Frederick Lefebvre 8min 2:53- 3:01
- ORNL – Buddy Bland 8min 3:02- 3:10
- Questions 5min 3:10- 3:15

UNIFY AND IMPROVE METHODOLOGY

- HPL and RandomAccess measurement methodologies are well established
- Green500 & TOP500 power-measurement methodology
  - Similar, but not identical methodologies
- Issues/concerns with power-measurement methodology
  - Variation in start/stop times as well as sampling rates
  - Node, rack or system level measurements
  - What to include in the measurement (e.g., integrated cooling)
- Current power measurement methodology is very flexible, but compromises consistency between submissions
- Proposal is to keep flexibility, but keep track of rules used and quality of power measurement

PROGRESS SINCE SC BOF

- Beta testing phase completed
  - ORNL, ANL, LRZ, University of Jaume, University of Tennessee
- Continued focus on improving power measurement methodology based on beta testing feedback
  - Idle time
  - Environmental conditions
- Power/Energy Measurement Methodology Document
  - http://www.green500.org/submissions
- Green500 soliciting L2 and L3 submissions

THE GREEN 500™

Submissions

You must login to submit an entry

Please use the 'Log in' link at the bottom of the page to log in or create a new account.

Submission deadline **Friday, June 14, 2013, 11:59:59 PM EDT**

NOTE: This form is for Level 1 submissions only! If you have a Level 2 or Level 3 submission please contact us directly to process your submission.

Share

Run Rules

Click the link below to download the latest Run Rules to find out what you need to do before you submit an entry.

Download the latest Run Rules (PDF)

EEHPC WG: Power Measurement Methodology

Click the link below to download the EEHPC WG Power Measurement Methodology to find out more about Level 2 and Level 3 measurements.

Download the EEHPC WG: Power Measurement Methodology Document (PDF)

Methodology Overview

Three Quality Levels

Level 1: basic measurement

Level 2: reasonable effort

Level 3: current best for HPC systems

ALL aspects for a quality level must be satisfied for a measurement to achieve that level

Four Aspects Required for each Level

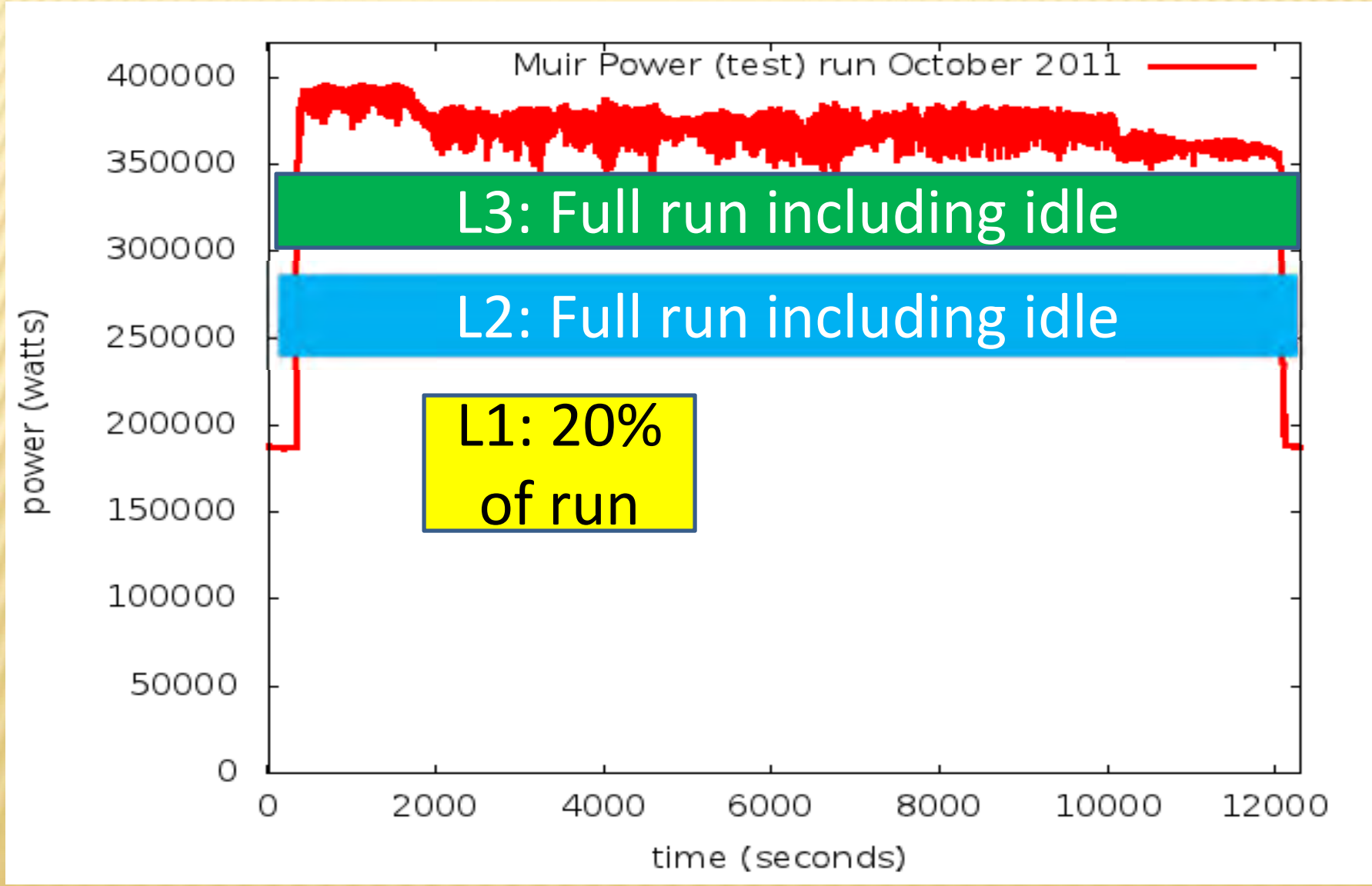
Aspect 1: frequency and time extent of measurements

Aspect 2: system fraction actually measured

Aspect 3: subsystems included

Aspect 4: power measurement location

Aspect 1: Time Extent



Aspect 1: Sampled Data Frequency

Level 3: (L3)

- "Continuously integrated" energy (≥ 120 samples per second)

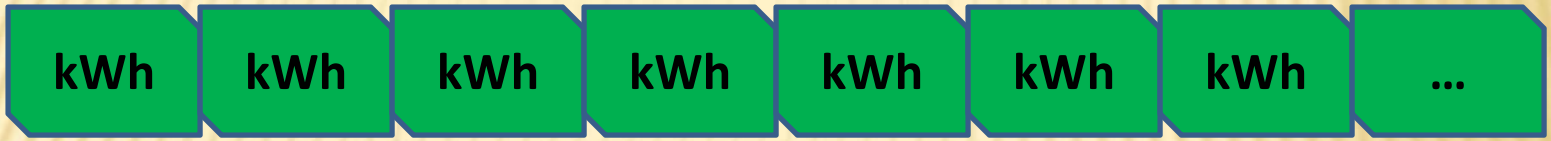
Level 1 and Level 2 (L1 and L2)

- Average power at least once per second

These are *sampling* rates. Data at this rate is typically not seen directly, it's internal to the device.

Aspect 1: Reported Data Requirements

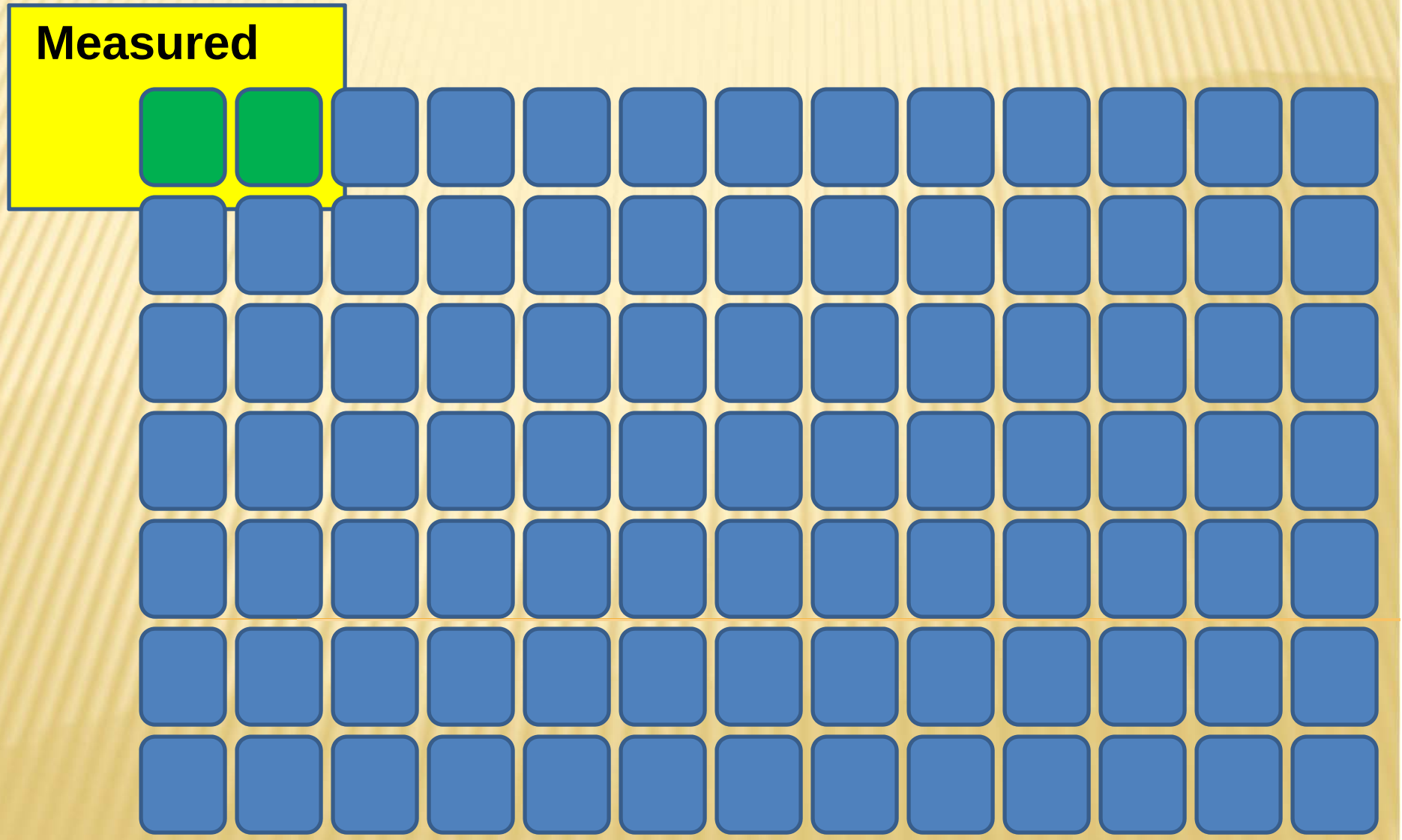
L3: at least 10 reported integrated energy values



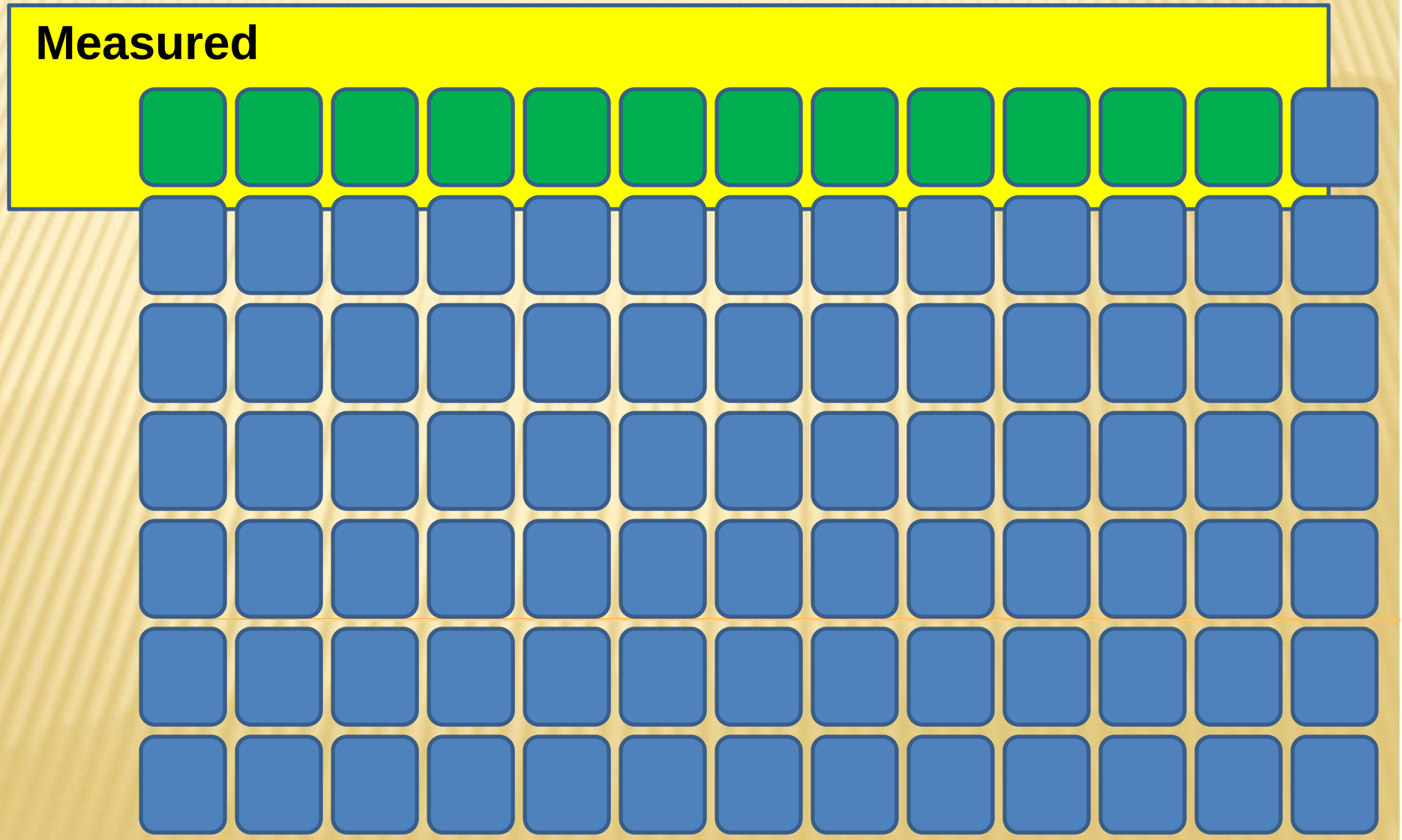
L2: at least 10 power averaged values

L1: at least one power averaged value

Aspect 2: Machine Fraction
L1: at least 1/64 or 1 kW

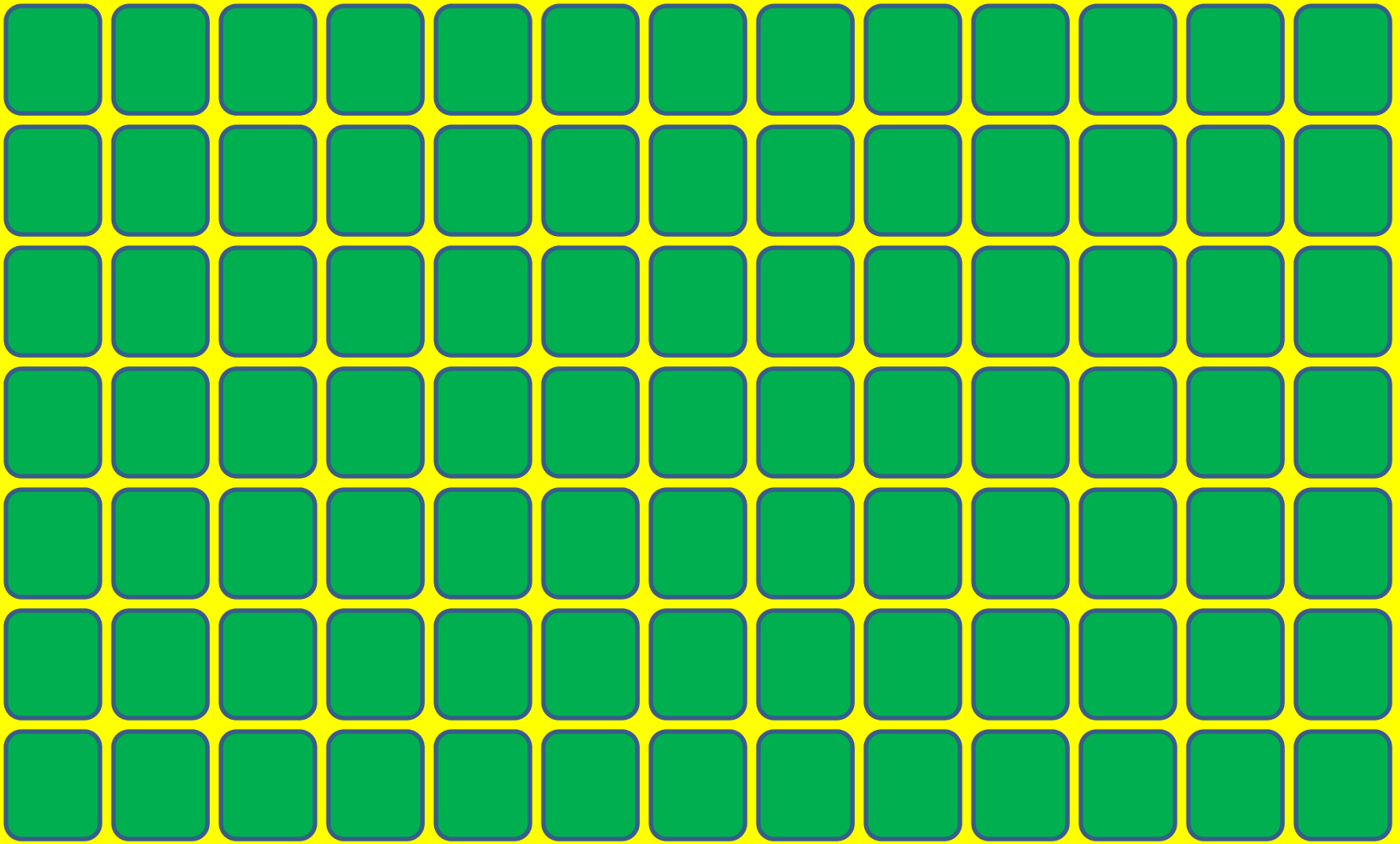


Aspect 2: Machine Fraction
L2: at least 1/8 or 10 kW



Aspect 2: Machine Fraction
L3: whole machine

Measured

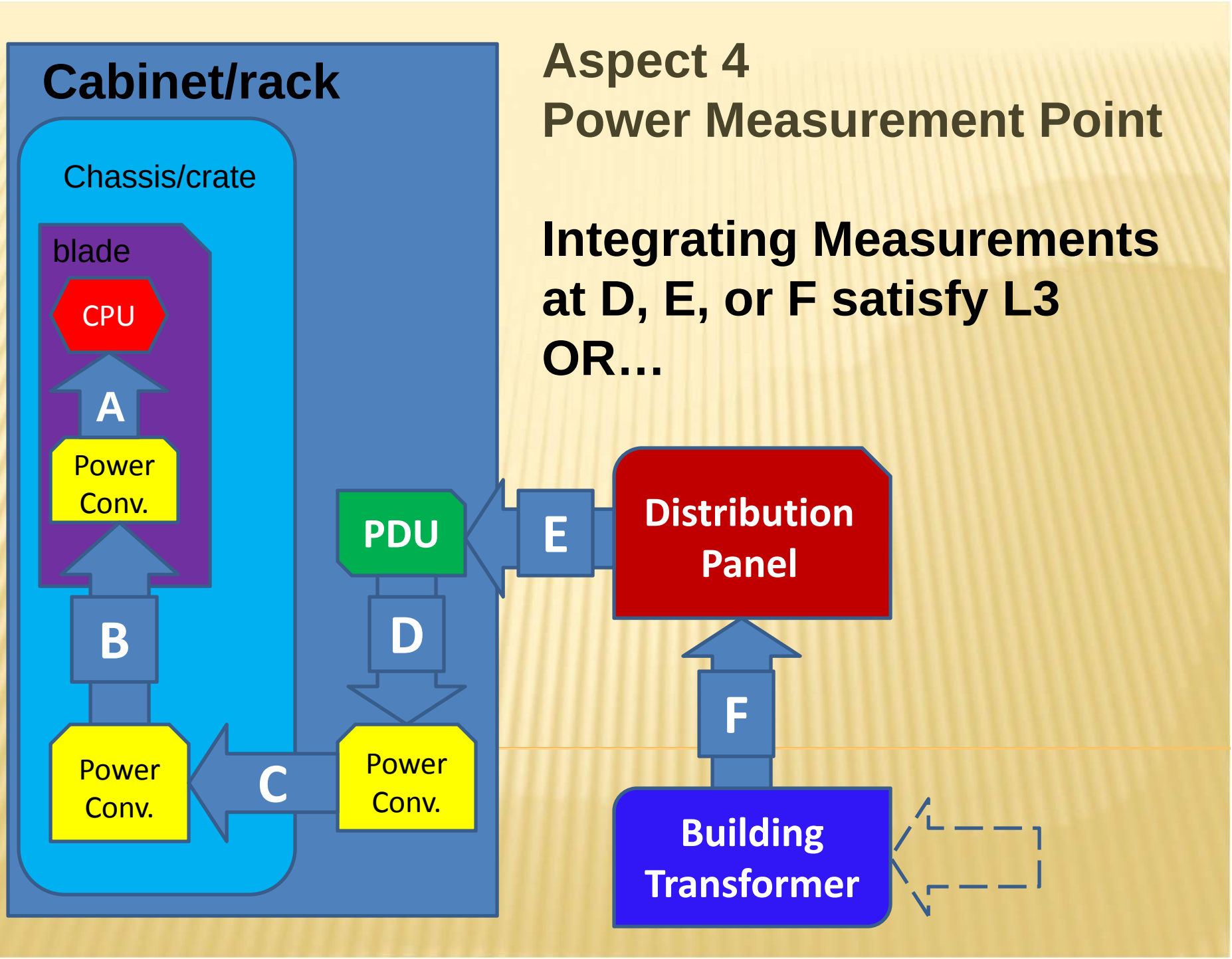


Aspect 3: Subsystem Inclusion

General philosophy: include all parts of computational system that participate in the workload

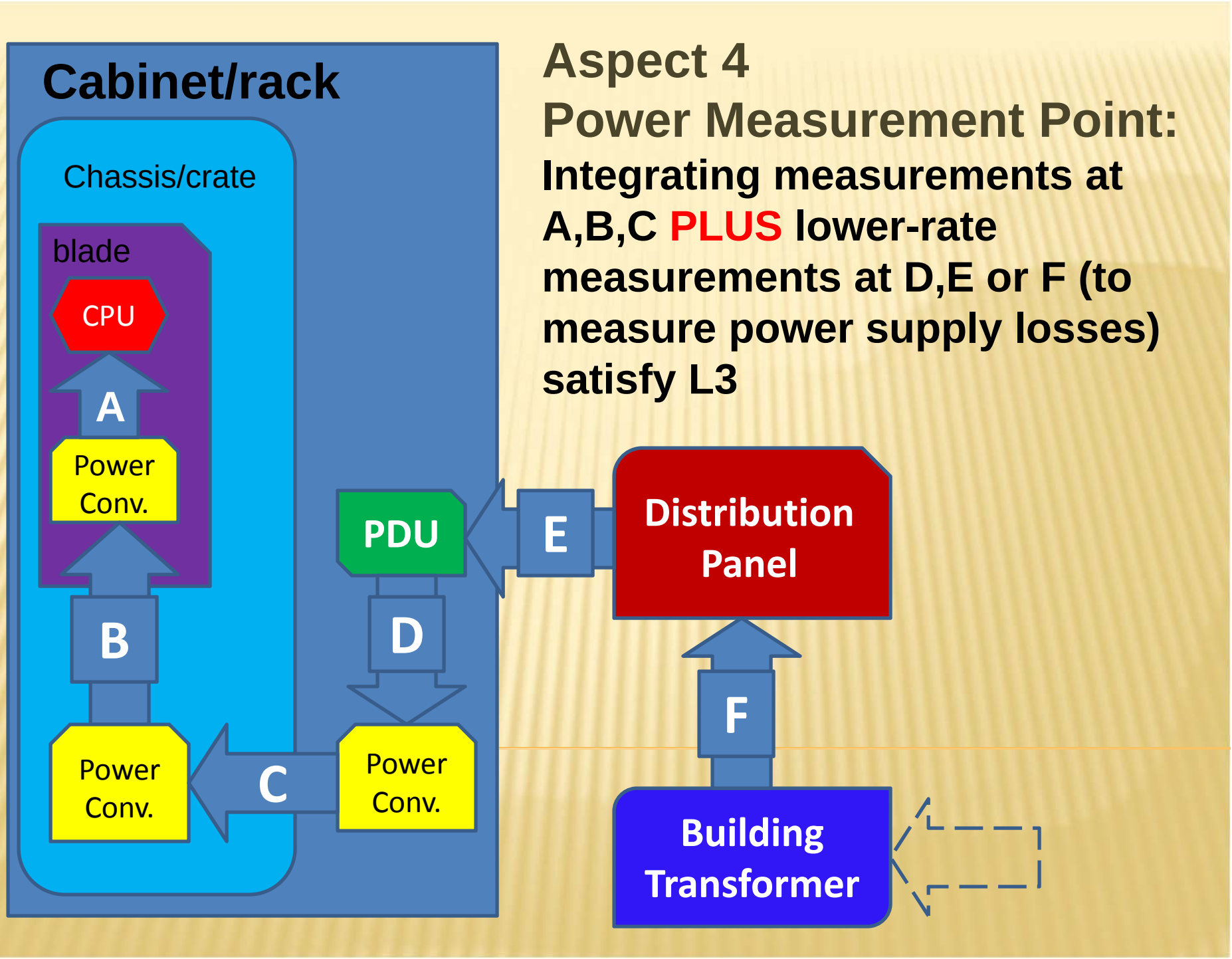
Must include:

- Processors, memory, cooling power internal to the machine (fans, etc.)
- Internal Interconnect network
- Login/compile nodes



Aspect 4
Power Measurement Point

Integrating Measurements at D, E, or F satisfy L3 OR…



Aspect 4
Power Measurement Point:

Integrating measurements at A,B,C PLUS lower-rate measurements at D,E or F (to measure power supply losses) satisfy L3

Levels 1-3 Summary Table

	Aspect 1	Aspect 2	Aspect 3	Aspect 4
L1	20% of run: 1 average power measurement	(larger of) 1/64 of machine or 1kW	[Y] Compute nodes [Y] Interconnect net [ ]Storage [ ]Storage Network [Y]Login/Head nodes	Measurement Location: [ ]A + D,E,F [ ]B + D,E,F [ ]C + D,E,F [ ]D [ ]E [ ]F
L2	100% of run: at least 10 average power measurements	(Larger of) 1/8 of machine or 10kW		
L3	100% of run: at least 10 running total energy measurements	Whole machine		

WHY WE ARE HERE

- To gather community momentum and support
- To solicit your feedback
- To ask you to participate by submitting measurements for Top500 and Green500 Lists
- To encourage higher quality measurements

Energy Efficient HPC Working Group

EE HPC WG

- Driving energy conservation measures and energy efficient design in high performance computing.

- Demonstrate leadership in energy efficiency as in computing performance.

- Forum for sharing of information (peer-to-peer exchange) and collective action.

http://eehpcwg.lbl.gov

Agenda

- Introduction – Tom Scogland 10min 2:15- 2:25
- LRZ – Torsten Wilde 8min 2:26- 2:34
- NCAR – Aaron Andersen 8min 2:35- 2:43
- CEA – Jean-Philippe Nomine 8min 2:44- 2:52
- Calcul Quebec – Frederick Lefebvre 8min 2:53- 3:01
- ORNL – Buddy Bland 8min 3:02- 3:10
- Questions 5min 3:10- 3:15

EXTRA SLIDES

WHY WE ARE HERE

- Context
  - Power consumption and facility costs of HPC are increasing.
- “Can only improve what you can measure”
- What is needed?
  - Converge on a common basis for:
    - METHODOLOGIES
    - WORKLOADS
    - METRICS

  for energy-efficient supercomputing, so we can make progress towards solutions.

AGREEMENT IN PRINCIPAL

- Collaboration between Top500, Green500, Green Grid and EE HPC WG
  - Discussions with Green.Graph500 as well
- Evaluate and improve methodology, metrics, and drive towards convergence on workloads
- Form a basis for evaluating energy efficiency of individual systems, product lines, architectures and vendors
- Target architecture design and procurement decision making process

PROPOSED METRIC GRANULARITY

- Measure behavior of key system components including compute, memory, interconnect fabric, storage and external I/O
  - Workloads and Metrics might address several components at the same time
- Phased implementation planned

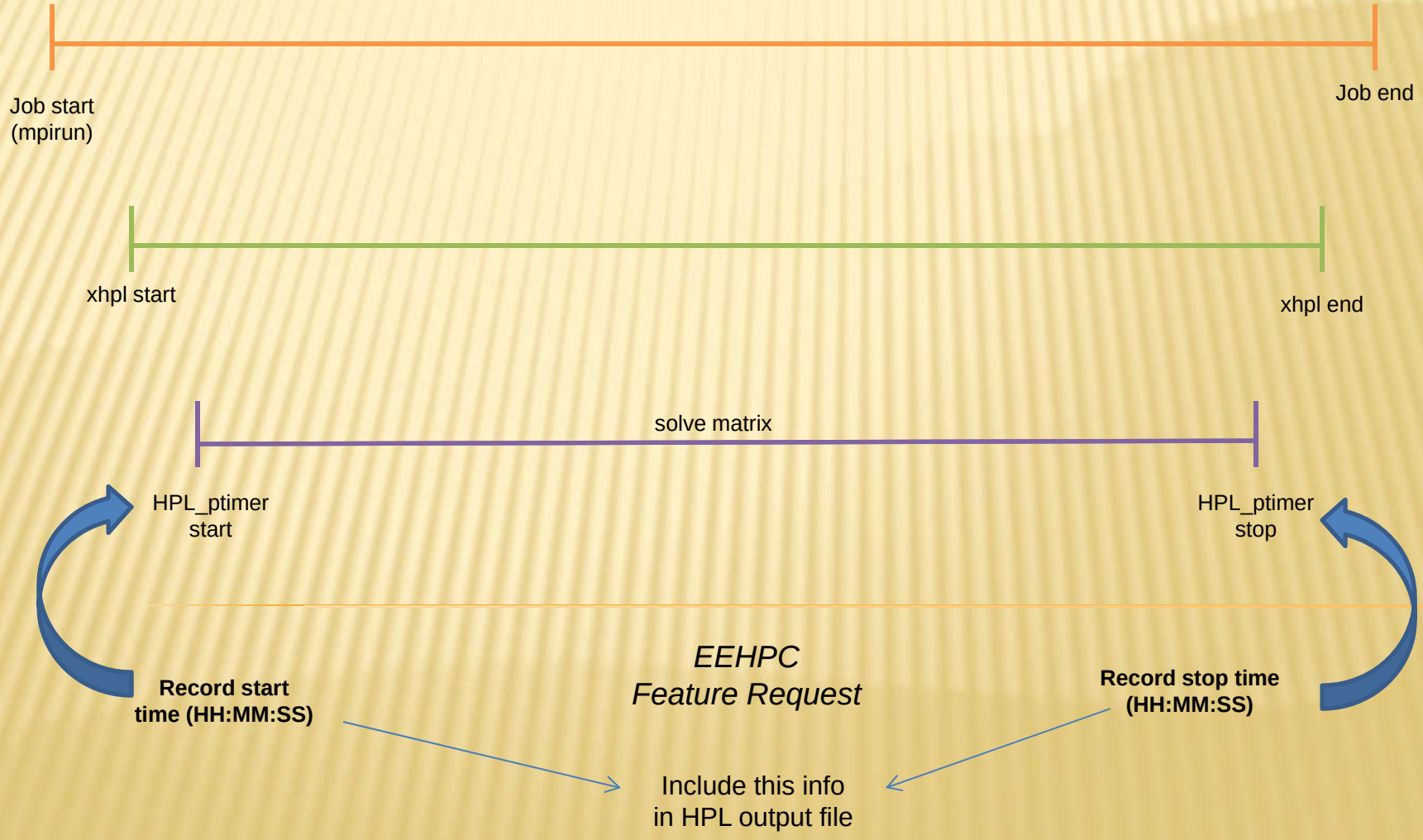
PROPOSED WORKLOADS

- Leverage well-established benchmarks
- Must exercise the HPC system to the fullest capability possible
- Use High Performance LINPACK (HPL) for exercising (mostly) compute sub-system
- Use RandomAccess (Giga Updates Per second or GUPs) for exercising memory sub-system (?)
- *Need to identify workloads for exercising other sub-systems*

ISSUES TO RESOLVE

- *Identify workloads for exercising other sub-systems; e.g., memory, storage, I/O*
- *Still need to decide upon exact metric*
  - *Classes of systems (e.g., Top50, Little500)*
  - *Multiple metrics or a single index*
  - *Energy and power measurements*
    - *Average power, total energy, max instantaneous power*

HPL Start/Stop Timestamps Needed to Identify Power Measurement Interval



Timestamps have been implemented

```
/*
 * Solve linear system
 */
   HPL_ptimer_boot(); (void) HPL_barrier( GRID->all_comm );
   time( &current_time_start );
   HPL_ptimer( 0 );
   HPL_pdgesv( GRID, ALGO, &mat );
   HPL_ptimer( 0 );
   time( &current_time_end );
```