SNIA

Education

# Green Storage and Energy Efficiency

Carlos Pratt, IBM

# SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.
  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

# Acknowledgement

- Initial tutorial
  - Erik Ridel (EMC) and Patrick Stanko (Oracle)
- Best foot forward
  - Jim Espy (EMC)
  - Herb Tanzer (Hp)
- In the technology area and validation of much of the work presented here
  - Alan Yoder (Huawei)
- Members of SNIA GSI and Green Technical Working Group
- Emerald™ Director
  - Dave Thiel

# Agenda

- SNIA green activities
  - Green Storage Initiative
  - Green Technical Working Group
  - Emerald™ Program
  - Why should storage consumers use the emerald program?
- Background and green storage
  - Revisit a basic storage unit
  - What influences green storage
  - Storage Power example
  - Introduce metrics
- Storage taxonomy
- the use of the Emerald™ Program
  - Basic configuration and requirements
  - Specification Requirements
  - Conditioning Test Phase
  - Active Test IO Profile
  - Ready Idle
  - Capacity Optimization Method Test
  - Valid Flow to generate data
- The Best Foot Forward
- Other Associations Green Storage Efforts

# Green Activities

- SNIA green activities
  - Green Storage Initiative
  - Green Technical Working Group
  - Emerald™ Program
  - Why should storage consumers use the emerald program?

# SNIA Green Activities
# Green Storage Initiative

## SNIA Green Storage Initiative (GSI)

- To conduct research on power and cooling issues confronting storage administrators
- Educate the vendor and user community about the importance of power efficiency in shared storage environments
- Leverage SNW and other SNIA and partner conference to focus attention on energy efficiency for networked storage infrastructures
- Provide input to the SNIA Green Storage TWG on requirements for green storage metrics and standards
- Provide external advocacy and support of SNIA Green TWG technical work

# SNIA Green Activities
# Green Technical Working Group

- Technical body working on green storage metrics and standards
- Gets direction from GSI
- Wrote the SNIA Emerald™ Power Efficiency Measurement Specification and related documents
- Supports the Emrald™
  - White papers
  - Tutorials
  - Training
- Works with regulatory agencies; i.e. EPA, on green storage specifications

# SNIA Emerald™ Program Overview



- Purpose
  - Provide open access to storage system power efficiency information using a well-defined testing procedure and additional information related to system power characteristics
  - The report data can help IT professionals make storage platform selections as part of an overall Green IT and Sustainability objective
  - Easily identifiable program logo
- Test procedure:
  - SNIA Emerald™ Power Efficiency Measurement Specification
- Public access and submittal is through the sniaemerald.com web site
  - No charge for access to test results, specifications or user guides
  - Submission of results is for a modest fee, discounted or waived for SNIA/GSI members
  - SNIA membership is not required to submit or to access test results
  - Voluntary, non-exclusionary, low cost program for manufacturers - Options to self-measure or third party measurement

# SNIA Emerald™ Program Overview - 2


- Process
  - Storage Vendors test their equipment and submit test results to the Emerald Program
  - Emerald Program publishes results on the sniaemerald.com web site
  - IT users (public) download results from the sniaemerald.com web site
  - Vendor gains right to use the SNIA Emerald™ logo in conjunction with tested products
- Legal protections
  - Terms of Use: conditions on use of test results agreed to by those downloading results
  - Terms of Submission: agreed to by vendor submitting test results
- Sign up for the mailing list: sniaemerald.com

# Why Should Storage Consumers Use the Emerald™ Program? – Overview 3

- SNIA Emerald™ Program seeks to
  - Provide a collection of standard metrics and data that allows IT architects to objectively compare a range of possible storage solutions
- SNIA Emerald™ Program
  - Enables users to select the mode of storage usage that accomplishes their work objectives with the lowest overall energy consumption
  - Drives vendor companies to innovate and compete in the development of energy efficient products as measured by the standard yardsticks

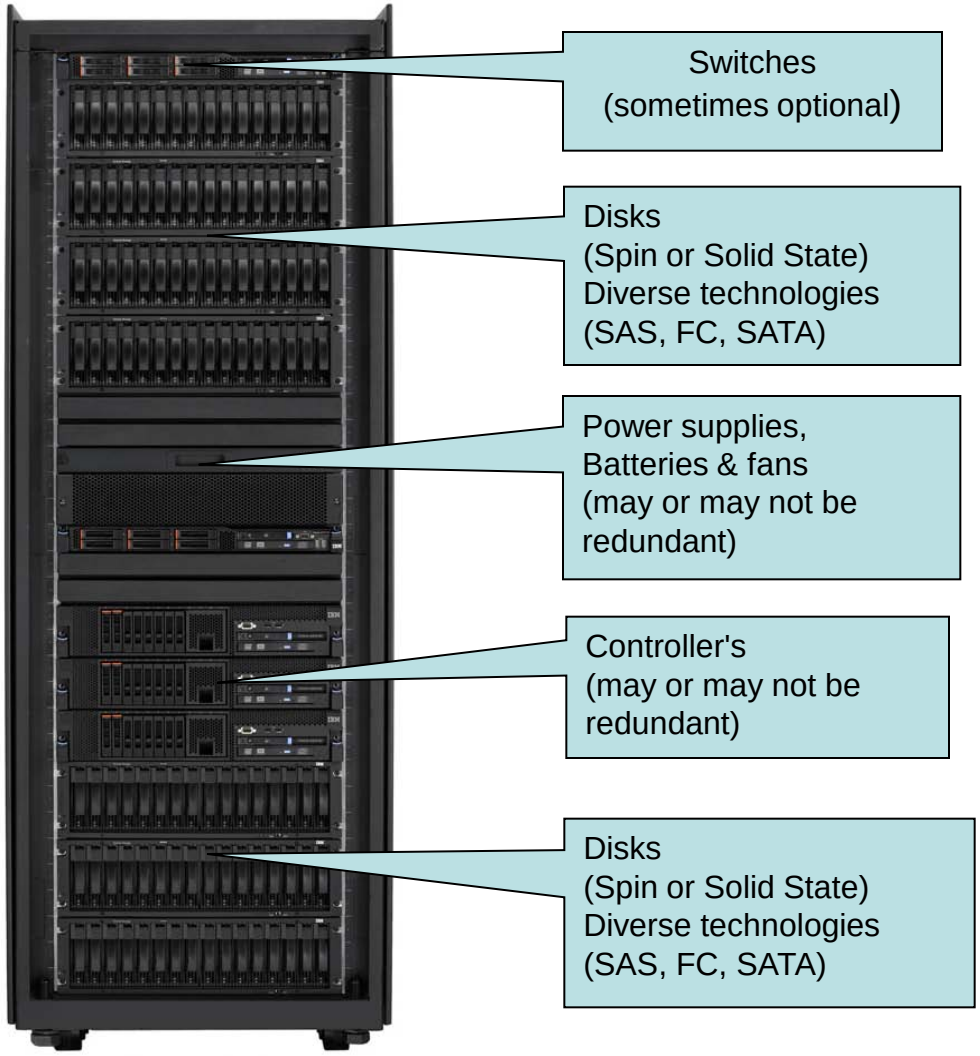
**The question is:**

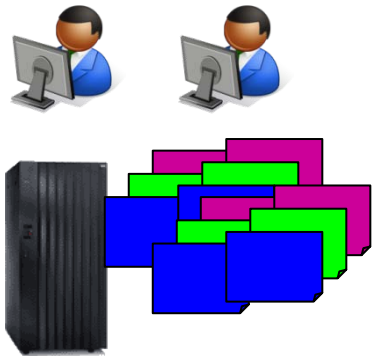
**WHERE IS THE ENERGY USED???**

# Background and Green Storage

- Revisit a basic storage unit anatomy and its use
- What influences green storage power consumption
- Storage power example
- Opportunities to make storage greener
  - (OK lets call it by its name … Use less Power)
- Introduce metrics
  - And the meaning for the IT managers?

# Basic Anatomy of a Disk Storage System and its Use



Users and Apps

Other:

Software (firmware & microcode)

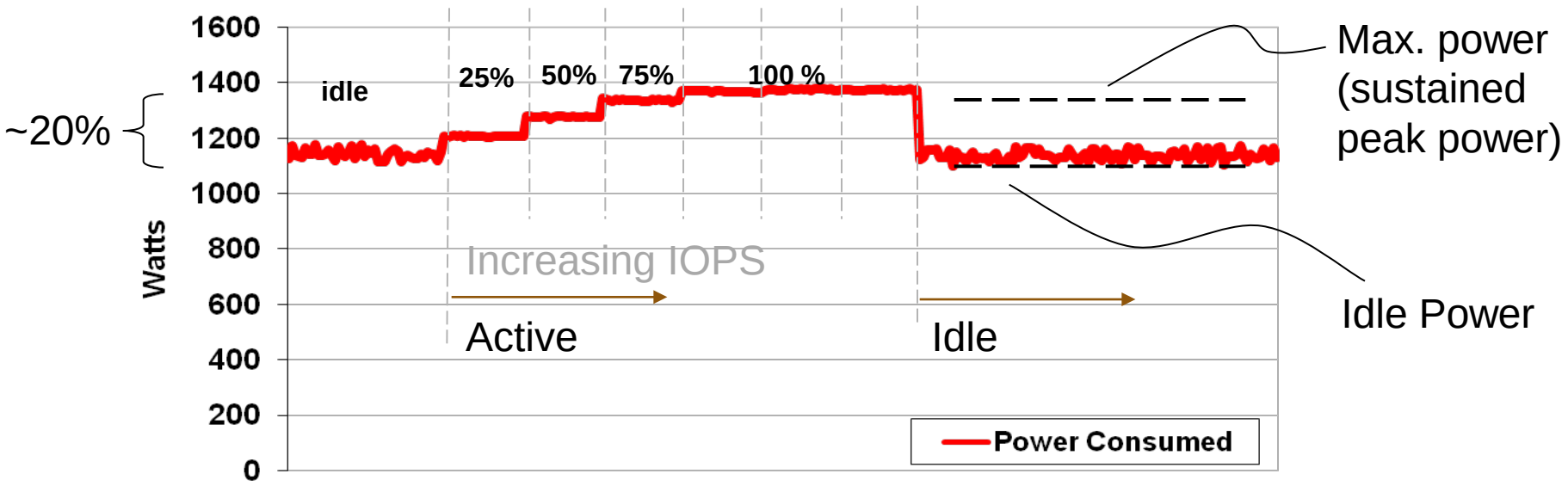
Hardware Design

Environment

# What Influences Power Consumption

- Storage capacity / usage efficiency
  - increasing data → larger capacity → more disks
  - redundant copies → magnify capacity needs
  - variability in usage and utilization → inefficient allocation of space
  - What is valuable data? What is the retention policy?
- Data transfer rate / access speed
  - high I/O bandwidth → higher rotational speed; striping across many drives
  - low access times → faster actuators; higher rotational speeds; caches
  - How fast and immediate must data be available? (time-to-data)
- Data integrity
  - 25% of "digital universe" is unique, but 75% are replicas / duplicates
  - partly to ensure data integrity and survivability; partly wasteful
- Data availability / system reliability
  - RAID uses extra drives, plus redundant power supplies, fans, controllers,
  - How valuable is data? How likely are failures? How fast must data be available?

# Storage Power Example




- Ideally, systems consume minimum power in all modes
  - Example system consumes **significant power in ready idle** (**80% of max**)
- % of time in Idle versus Active depends on storage type, application and workloads; available optimizations will vary
  - Power itself is only one part of the story it must be reflected as a metric
- Power consumed is not linearly proportional to workload

# Opportunities to Make Storage Green(er)

- Environment
  - Higher system tolerance to high/lower temperatures and humidity
  - In line with cold and hot aisles designs on new data centers
- Improve usage efficiency
  - De-duplication and compression
  - Thin provisioning
- Minimize energy consumption
  - Improved component designs – high-efficiency power supplies, advanced & flexible storage devices
  - Variants of MAID – idle and spin-down
- New technologies
  - Solid state storage
  - Alternative + hybrid system designs (opportunity to rethink)

must be driven by metrics / standards / guidelines

# Green Storage Metrics

- SNIA recommended metrics
  - Capacity metric (ready-idle)
    - Relates the power of the system to its total storage raw capacity. It is reported as GB/watt (or TB/watt)
    - Power required to store and protect the data
  - Workload metric (Active – Also known as transactional)
    - Relates the power of the system to the maximum possible IOPS generated by a specific random stress load. It is reported as IOPS/watt
    - Power required to randomly supply data to and from a host
  - Bandwidth metric (Active – Also known as streaming)
    - Relates the power drown by the system to the maximum possible MBPS generated by a specific sequential stress load. It is reported as MBPS/watt
    - Power required to stream data to and from a host

# What is the Ready Idle Metric and what it Means to an IT Manager?

- Depending on the systems and their usage their energy usage may be evaluated according to:
  - Is the system idle time at least 12 hours or more a day?
    - You should be interested in the power required to store the data
    - **capacity metric (GB/Watt)** may be your best indicator on how energy efficient your system is
    - The larger this number is the less watts are used to energize the total storage of your system

# What are the Active Metrics and what it Means to an IT Manger?

Education SNIA

- Depending on the systems and their usage their energy usage may be evaluated according to:
  - For systems running more than 12 hours a day
    - You should be interested in the Power to move the data onto and off the storage system
    - Is your load predominantly sequential?
      - **Bandwidth metric (MBS/Watt)** will help you to determine how effective is your power use. The larger this number is, the more data the system is pushing per watt
    - Is your load predominantly random?
      - **Workload metric (IOPS/Watt)** will help you determine how effective is your power use. The larger this number is the system is provides more operations per watt.
    - Independently on how long the system is idle it is always good to know what is your capacity per watt ratio

# Storage Taxonomy

- Taxonomy
  - 1. the science or technique of classification.
  - 2. a classification into ordered categories: a proposed taxonomy of educational objectives.
  - 3. Biology . the science dealing with the description, identification, naming, and classification of organisms.

Source:

Dictionary.com

# Why Have a Storage Taxonomy

- Need a taxonomy to enable fair comparisons among similar storage products
  - e.g. for motor vehicles – motorcycles, cars, trucks
- Similar green metrics may apply to all product categories, but different values establish best-in-class
- Unique considerations apply to special categories
  - e.g. amphibious cars, skid steer loaders, tanks
- Clear taxonomy will simplify comparisons and aid regulatory efforts
- SNIA Storage Taxonomy is defined in SNIA Emerald™ Power Efficiency Measurement Specification

# Taxonomy – Categories

| Attribute | Category | | | | | |
|---|---|---|---|---|---|---|
| | Online | Near Online | Removable Media Library | Virtual Media Library | Adjunct Product | Interconnect Element |
| Access Pattern | Random/ Sequential | Random/ Sequential | Sequential | Sequential | | |
| MaxTTFD (t) | t < 80 ms | t > 80 ms | t > 80 ms<br>t < 5 min | t < 80 ms | t < 80 ms | t < 80 ms |
| User Accessible Data | Required | Required | Required | Required | Prohibited | Prohibited |

- Six categories, covering most storage industry products

# Taxonomy – Categories

| Category → / Level | Online | Near Online | Removable Media Library | Virtual Media Library |
|---|---|---|---|---|
| Consumer/ Component | Online 1 | Near Online 1 | Removable 1 | Virtual 1 |
| Low-end | Online 2 | Near Online 2 | Removable 2 | Virtual 2 |
| Mid-range | Online 3 | Near Online 3 | Removable 3 | Virtual 3 |
| | Online 4 | | | |
| High-end | Online 5 | Near Online 5 | Removable 5 | Virtual 5 |
| Mainframe | Online 6 | Near Online 6 | Removable 6 | Virtual 6 |

| Adjunct Product | Interconnect Element |
|---|---|
| | |
| | |
| | |
| | |
| | |

- 23 total "buckets" covering the breadth of the industry

# Taxonomy – Online

- Most common storage systems

| Attribute | Classification | | | | | |
|---|---|---|---|---|---|---|
| | Online 1 | Online 2 | Online 3 | Online 4 | Online 5 | Online 6 |
| Access Pattern | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential |
| MaxTTFD (t) | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms | t < 80 ms |
| User-Accessible Data | Required | Required | Required | Required | Required | Required |
| Consumer/ Component | Yes | No | No | No | No | No |
| Connectivity | Not specified | Connected to single or multiple hosts | Network-connected | Network-connected | Network-connected | Network-connected |
| Maximum Configuration | ≥1 | ≥ 4 | ≥ 12 | > 100 | >400 | >400 |
| Integrated Storage Controller | Optional | Optional | Required | Required | Required | Required |
| Storage Protection | Optional | Optional | Required | Required | Required | Required |
| No SPOF | Optional | Optional | Optional | Required | Required | Required |
| Non-Disruptive Serviceability | Optional | Optional | Optional | Optional | Required | Required |
| FBA/CKD Support | Optional | Optional | Optional | Optional | Optional | Required |

# Taxonomy – Near Online

| Attribute | Classification | | | | | |
|---|---|---|---|---|---|---|
| | Near Online 1 | Near Online 2 | Near Online 3 | Near Online 4 | Near Online 5 | Near Online 6 |
| Access Pattern | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential | Random/ Sequential |
| MaxTTFD (t) | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms | t > 80 ms |
| User-Accessible Data | Required | Required | Required | Required | Required | Required |
| Consumer/ Component | Yes | No | No | No | No | No |
| Connectivity | Not specified | Connected to single or multiple hosts | Network-connected | Network-connected | Network-connected | Network-connected |
| Maximum Configuration | ≥1 | ≥ 4 | ≥ 12 | > 100 | >400 | >400 |
| Integrated Storage Controller | Optional | Optional | Required | Required | Required | Required |
| Storage Protection | Optional | Optional | Required | Required | Required | Required |
| No SPOF | Optional | Optional | Optional | Required | Required | Required |
| Non-Disruptive Serviceability | Optional | Optional | Optional | Optional | Required | Required |
| FBA/CKD Support | Optional | Optional | Optional | Optional | Optional | Required |

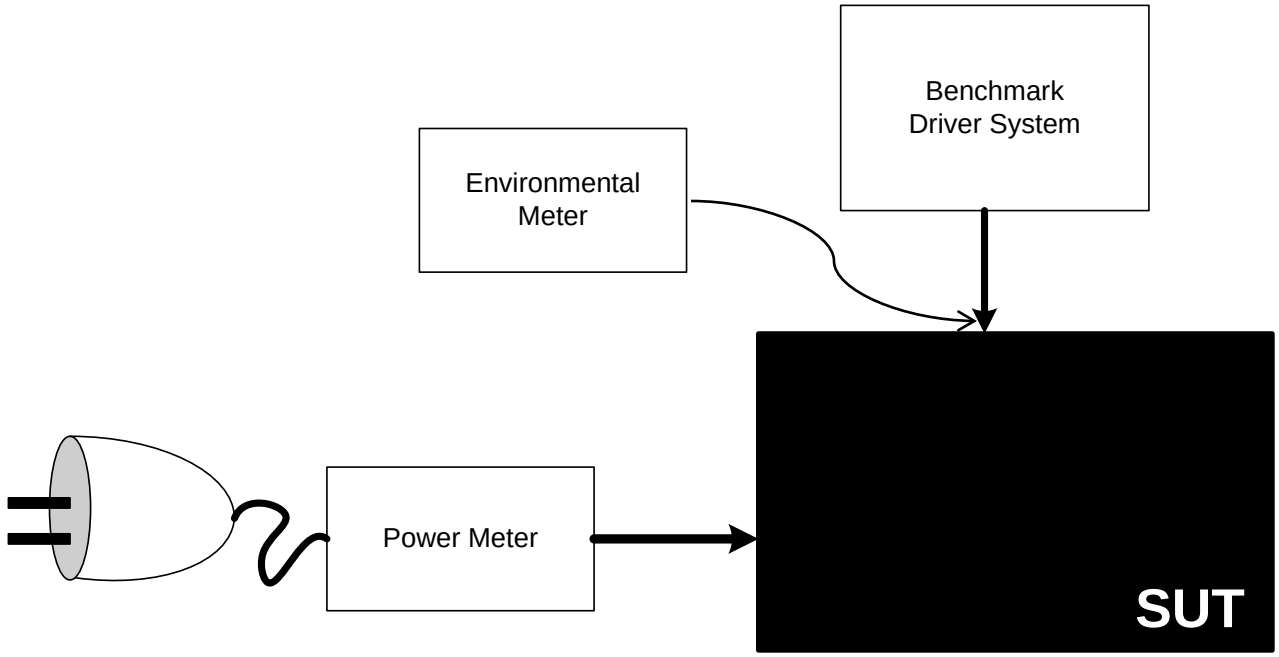
For more taxonomy definitions please consult the emerald spec.

# The use of the emerald program ™

- Basic configuration and requirements
- Specification Requirements
- Conditioning Test Phase
- Active Test IO Profile
- Ready Idle
- Capacity Optimization Method Test
- Valid Flow to generate data

# Test Configuration and Requirements

- Basic configuration
  - Not allowed to change configuration or tune parameters during test phases



# Specification Requirements

- Test setup requirements
  - Input voltage – standard world voltages
  - Environmental – standard datacenter conditions
  - Benchmark driver – VdBench, IOMeter, required conditions
  - Meters
    - Power – accuracy level
    - Temperature- accuracy level
- Equations
  - Average response time and power
  - Periodic power efficiency
  - Metric stability – 10 point rolling average
  - Time interval of 1 minute or a specified measurement interval

Some important changes to the requirements are expected on version 2 of the Emerald Spec

# SUT Conditioning Test Phase

- Intended to provide a uniform initial condition for subsequent measurements
- Demonstrate the SUT's ability to process IO requests
- Assure that each storage device in the SUT is fully operational and capable of satisfying any supported request
- Achieve typical operational temperature
- Each taxonomy category will have different measurement interval requirement to demonstrate stability

# Active Test IO Profiles

| IO Profile | IO Size (KiB) | Read/Write Percentage | IO Intensity | Transfer Alignment (KiB) | Access Pattern |
|---|---|---|---|---|---|
| Mixed Workload 1 (i=MW1) | 8 | 70/30 | 100 | 8 | Random |
| Mixed Workload 2 ( i=MW2) | 8 | 70/30 | 25 | 8 | Random |
| Random Write ( i=RW) | 8 | 0/100 | 100 | 8 | Random |
| Random Read ( i=RR) | 8 | 100/0 | 100 | 8 | Random |
| Sequential Write ( i=SW) | 256 | 0/100 | 100 | 256 | Sequential |
| Sequential Read ( i=SR) | 256 | 100/0 | 100 | 256 | Sequential |

- All or some of the IO profiles are used by the defined taxonomy categories
  - Drive enough IOs to reach the required response time or through-put specified in the measurement specification
  - The 25 IO intensity is 25% of the IO defined for MW1

# SUT Active Test Phase

- IO profiles used by taxonomy category
  - Online and Near-Online use all six IO profiles
  - Removable Media and VML use only the sequential IO profiles
- Run as an uninterrupted sequence of workloads
  - Specification defines the order to be run for each taxonomy category

# SUT Ready Idle Test Phase

- Defined as storage systems and components that are configured, powered up, connected to one or more hosts and capable of satisfying externally-initiated, application-level initiated IO requests within normal response time constraints, but no such IO requests are being submitted.
- Average power measured in the measurement window
- No external IO given by the host
- Can perform any IO within the taxonomy required response time interval

# SUT Capacity Optimization Method Test Phase

- Heuristic tests
  - Delta snapshots
  - Thin provisioning
  - Data de-duplication
  - Parity RAID
  - Compression
- Run after ready idle test phase
- C program generated by SNIA
  - Download from sourceforge.net/projects/sniadeduptest
  - Used for de-duplication and compression
- Taxonomy dependent

# Metrics

- Active (Primary)
  - Ratio of operations rate over average power for the same measurement interval
    - $EP_{MW1}$ (IOP/S/W) of the 70% mixed workload at maximum response time
    - $EP_{MW2}$ (IOP/S/W) of the 25% of the IO used in MW1
    - $EP_{RR1}$ (IOP/S/W) of the random read workload at maximum response time
    - $EP_{RW1}$ (IOP/S/W) of the random write workload at maximum response time
    - $EP_{SR1}$ (MiB/S/W) of the sequential read workload at maximum throughput
    - $EP_{SW1}$ (MiB/S/W) of the sequential write workload at maximum throughput
  - Energy

| IO Profile | IO Size (KiB) | Read/Write Percentage | IO Intensity | Transfer Alignment (KiB) | Access Pattern |
|---|---|---|---|---|---|
| Mixed Workload 1 (i=MW1) | 8 | 70/30 | 100 | 8 | Random |
| Mixed Workload 2 ( i=MW2) | 8 | 70/30 | 25 | 8 | Random |
| Random Write ( i=RW) | 8 | 0/100 | 100 | 8 | Random |
| Random Read ( i=RR) | 8 | 100/0 | 100 | 8 | Random |
| Sequential Write ( i=SW) | 256 | 0/100 | 100 | 256 | Sequential |
| Sequential Read ( i=SR) | 256 | 100/0 | 100 | 256 | Sequential |

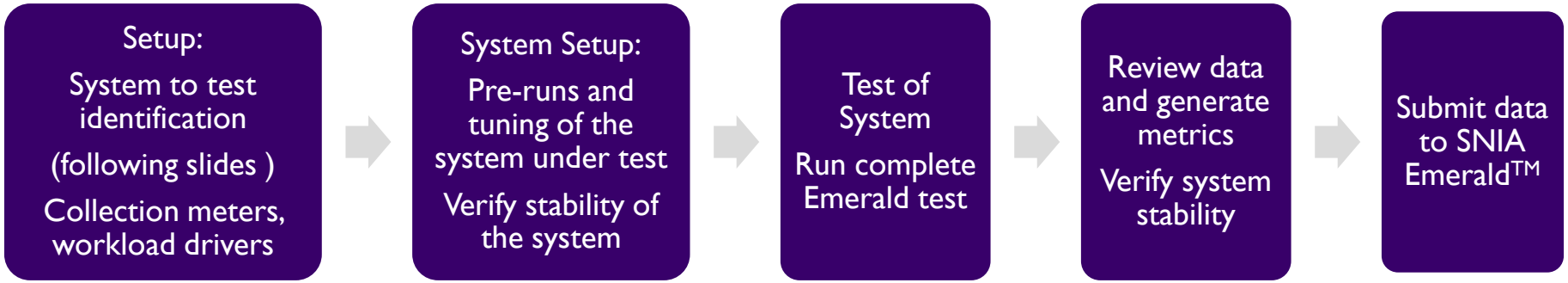
# Metrics (Continued)

- Ready Idle (Primary)
  - Ratio of raw capacity over average power measured in the defined measurement window (GB/W)
- Capacity Optimization (Secondary)
  - A yes/no for each Capacity Optimization Method tested
  - Do not have to test all COMs but if vendor declares to have a COM it must be tested and on during active test phase



© 2012 Storage Networking Industry Association. All Rights Reserved.

# Flow Needed for Valid Emerald Measurement




- General timeline
  - Tune the system
  - A day to run test
  - A day to generate the required data and review it
  - A few hours to submit the data

# Best Foot Forward (BFF) (Sweet Spot testing)

- Test configuration challenges
- Best Foot Forward Approach 1
  - Scale up system
- Best Foot Forward Approach 2
  - Scale out system
- Best Foot Forward Approach 3
  - Rough approximation

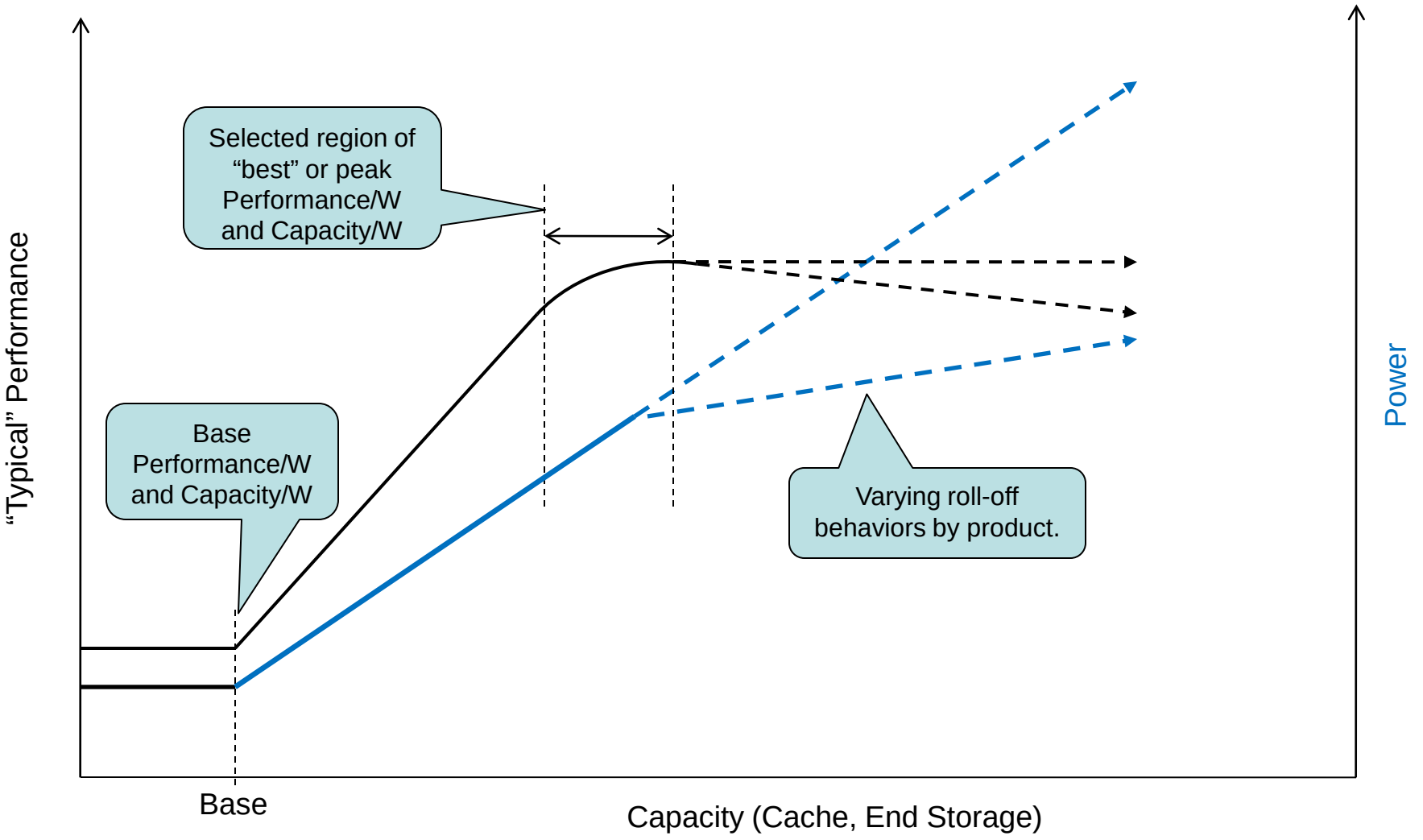
# Test Configuration Challenges

- Wide Spectrum of Storage-Oriented Products
  - Created a taxonomy to narrow scope
  - Categories: On-Line, Near-Line, etc.
  - Classifications: Further granularity of each Category
- Still too Broad in Scope
  - Vendors may have multiple products in a particular Category/Classification
  - Each product may have many configuration variables
- Requirement/Challenge: Select Appropriate Test Configurations
  - Comprehensive and usable results for customer
  - Minimized, lower cost, but effective testing methods for vendor

# Best Foot Forward Approach - 1

- BFF Looks Holistically at Storage System Product/Family
  - Allows vendor to select and test one product/family configuration
    - Or more if desired
  - At operating points near the Measurement Spec metric peak values
    - I.e. the "sweet spot"
  - Results reasonably representative of the entire family
    - Easier and less expensive for the vendor
    - Simple and understandable results for the potential customer
- Scale Up Example on Following Slide
  - Based on notion that Measurement Spec active metrics have peak values
  - Peaks typically located at points well below maximum configurations

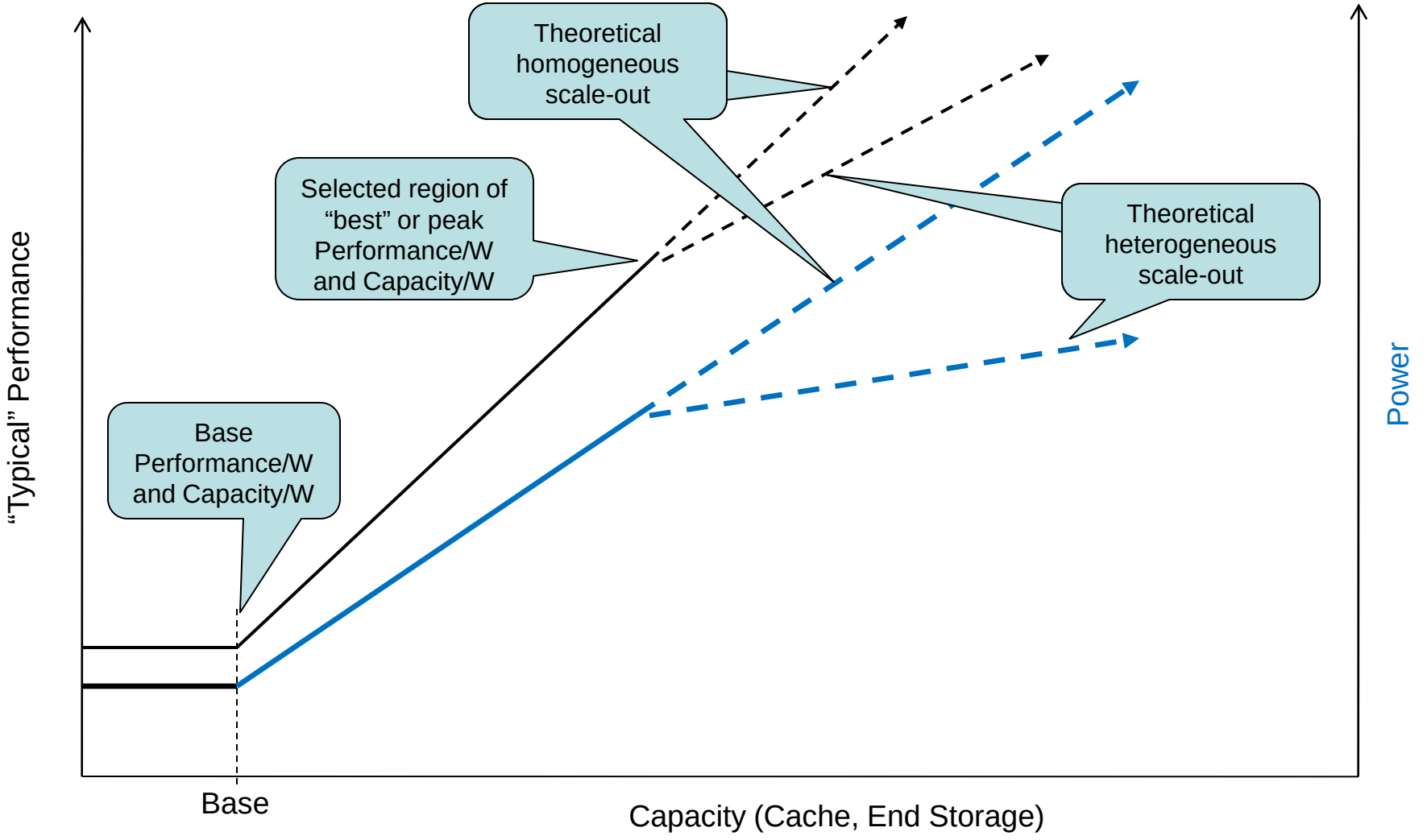
# Best Foot Forward Approach Scale-Up System



# Best Foot Forward Approach - 2

- Previous Slide is a Rough Approximation
  - Capacity increases are actually more stepwise
  - Performance roll-off can vary by product
    - Dashed lines attempt to show one (of possibly many) changes due to different storage technology tiers, e.g. scaling capacity w/large SATA drives
  - Regardless, example depicts a smaller test configuration
- What About Other Test Points?
  - Could also test at base (entry point) but not required
  - Key is no requirement to test beyond the peak point
- Scale Out Example on Following Slide
  - What if there is no clearly discernable peak?

# Best Foot Forward Approach Scale-Out System



# Best Foot Forward Approach - 3

- Again a Rough Approximation
  - Capacity increases are actually more stepwise
  - Dashed lines attempt to show one (of possibly many) changes due homogeneous vs heterogeneous scale-out configurations
  - Can still select a smaller test configuration

# Summary and Next Steps

- Given Known Taxonomy Category and Classification
  - Vendor determines one or more family representative configurations
  - Vendor locates Measurement Spec active metric peak points
  - Tests are performed on this reduced configuration (set)
    - Note: For smaller systems, the BFF may in fact be the maximum configuration
- Where is the Performance/W Peak?
  - Depends on numerical increase of numerator vs denominator with capacity
  - If numerator initially increases more than denominator, a clear peak
  - Else it becomes harder – Just pick a point before it rolls off?
  - Selection of the peak point(s) is the subject of the next slides

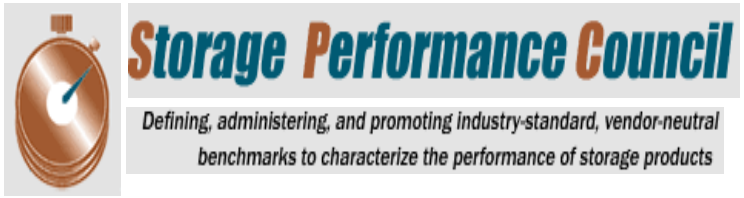
# Other associations work on storage efficiency

- SPC
- Green Grid

# Other Associations Green Storage Efforts

Education SNIA

- SPC




  - Storage Performance Council mainly oriented to disk subsystems was the first industry association to add power to their benchmark

- The Green Grid


  - Data Center Maturity Model
  - Design guide in progress
  - Working on a usage metrics
    - Measure the IOP/s/W, MiBs/W, GB/W in the datacenter

# Attribution & Feedback

The SNIA Education Committee would like to thank the following individuals for their contributions to this Tutorial.

## Authorship History

**Between 2009 and 2011**
**Erik Ridel**
**Carlos Pratt**
**Patrick Stanko**
**Patrick Chu**
**Matthew Brisse**
**David Reinsel**
**Edgar St.Pierre**
**Alan Yoder**
**Wayne Adams**
**SW Worth**

**Updates:**
**Carlos Pratt/August – September 2012**

## Additional Contributors

**Herb Tanzer**
**David Thiel**
**Jim Espy**

**ALL SNIA GSI and TWG members**

*Please send any questions or comments regarding this SNIA Tutorial to*
***tracktutorials@snia.org***

# SO, Where is the power?

Thank you