

12-9-2022

Design, development and evaluation of the ruggedized edge computing node (RECON)

Sahil Girin Patel

Mississippi State University, sahilp94@gmail.com

Follow this and additional works at: <https://scholarsjunction.msstate.edu/td>



Part of the [Hardware Systems Commons](#), [Heat Transfer, Combustion Commons](#), and the [Other Mechanical Engineering Commons](#)

Recommended Citation

Patel, Sahil Girin, "Design, development and evaluation of the ruggedized edge computing node (RECON)" (2022). *Theses and Dissertations*. 5695.
<https://scholarsjunction.msstate.edu/td/5695>

This Graduate Thesis - Open Access is brought to you for free and open access by the Theses and Dissertations at Scholars Junction. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.

Design, development and evaluation of the ruggedized edge computing node (RECON)

By

Sahil Girin Patel

Approved by:

HeeJin Cho (Major Professor)

Shanti Bhushan

Prashant Singh

Tonya W. Stone (Graduate Coordinator)

Jason M. Keith (Dean, Bagley College of Engineering)

A Thesis

Submitted to the Faculty of

Mississippi State University

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in Mechanical Engineering

in the Department of Mechanical Engineering

Mississippi State, Mississippi

December 2022

Copyright by
Sahil Girin Patel
2022

Name: Sahil Girin Patel

Date of Degree: December 9, 2022

Institution: Mississippi State University

Major Field: Mechanical Engineering

Major Professor: HeeJin Cho

Title of Study: Design, development and evaluation of the ruggedized edge computing node (RECON)

Pages in Study: 161

Candidate for Degree of Master of Science

The increased quality and quantity of sensors provide an ever-increasing capability to collect large quantities of high-quality data in the field. Research devoted to translating that data is progressing rapidly; however, translating field data into usable information can require high performance computing capabilities. While high performance computing (HPC) resources are available in centralized facilities, bandwidth, latency, security and other limitations inherent to edge location in field sensor applications may prevent HPC resources from being used in a timely fashion necessary for potential United States Army Corps of Engineers (USACE) field applications. To address these limitations, the design requirements for RECON are established and derived from a review of edge computing, in order to develop and evaluate a novel high-power, field-deployable HPC platform capable of operating in austere environments at the edge.

DEDICATION

This work is dedicated to my parents, Pratibha and Girin Patel, whose support, love, and dedication foundationally enable my pursuits and achievements.

ACKNOWLEDGEMENTS

I thank the US Army Corps of Engineers Engineer Research and Development Center for providing the foundation and environment for this effort, along with Dan Eng, Jonathan Boone, and Ivan Beckman for their assistance in propelling this research forward.

I additionally thank Dr. Rajeev Agrawal, Dr. Ruth Cheng, Ross Glandon, Anthony Lam, and Reed Williams for organizing and applying the high performance computing aspects of this project.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
I. INTRODUCTION	1
1.1 Background Motivations	1
1.2 Work Overview	3
1.3 Organization	4
II. BACKGROUND	5
2.1 Edge and Decentralized Networks	5
2.2 Edge Computing Motivations and Benefits	12
2.2.1 Latency	12
2.2.2 Communications Transmission Cost	13
2.2.3 Smart Data Management	15
2.2.4 Edge Device Processing and Energy Management	15
2.3 Edge Computing Challenges	16
2.3.1 Device Inflexibility	17
2.3.2 Appropriate Processing Matching	17
2.3.3 Edge Node Seamless Discoverability, Scalability	18
2.3.4 Seamless Operation Over Geospatial and Network Boundaries	19
2.3.5 Edge Component Location	20
2.3.6 Security	21
2.4 Edge Motivation and Challenges Conclusion	22
2.5 Edge Requirements for Computing	23
2.5.1 Cloud Infrastructure Edge Conversion Configuration	23
2.5.2 Small-Scale Edge Computing Configuration	24
2.5.3 Edge Computing Configuration Comparison	25
2.5.4 Edge Environment Requirements	26
2.5.5 Edge Configuration and Requirements Conclusion	27
2.6 Engineering Mission Objectives Background	28

2.7	USACE Mission Set Background	32
2.7.1	Military Environment Edge Computing	34
2.7.2	Disaster Environment Edge Computing	35
2.7.3	Mission Set Conclusion	37
2.8	Deployable Edge Computing Background	37
2.9	RECON Established Requirements	40
2.9.1	Objective	40
2.9.2	Requirements	41
2.9.3	Research	41
III.	CONCEPT DESIGN DEVELOPMENT	42
3.1	Introduction	42
3.2	Approach Overview	44
3.3	Computing Component Selection	45
3.3.1	Component Objectives	45
3.3.2	Computing Components	46
3.3.2.1	NVIDIA DGX-1	46
3.3.2.2	Dell Precision 7920 Rack Workstation	48
3.3.2.3	DDN SFA220NV	50
3.3.2.4	Mellanox InfiniBand SB7700	51
3.3.2.5	Raritan IX7 Power Controller	52
3.3.2.6	Computing Pipeline	52
3.4	Component Ruggedization	52
3.4.1	Objectives	52
3.4.2	Requirements	53
3.4.3	Approach	53
3.4.4	Medium Duty Polymer Rack Mount Cases	54
3.4.5	Heavy Duty Aluminum Rack Mount Cases	55
3.4.6	Ruggedization Selection	56
3.4.7	Ruggedization Conclusion	61
3.5	Cooling System Selection and Concept Integration	63
3.5.1	Cooling System Requirements	63
3.5.2	Approach	65
3.5.3	Integrated Coolers	65
3.5.4	Portable Coolers	66
3.5.5	Portable Cooler Concept	67
3.5.6	In-Row Server Coolers	68
3.5.7	Split Server Coolers	70
3.5.8	Split Server Cooler Concept	71
3.5.9	Cooling System Selection	73
3.6	Framing Material Selection	77
3.6.1	Approach	77
3.6.2	Framing Hardware Selection	78
3.6.3	Duct Material Selection	80
3.7	Concept Design	81

IV.	PROTOTYPE 1 DESIGN AND EVALUATION.....	84
4.1	Overview	84
4.2	Design Objectives.....	84
4.3	Design Response Overview.....	85
4.3.1	Framing Hardware Design	89
4.3.1.1	Base Chassis and Hardware.....	90
4.3.1.2	Vertical Framing.....	92
4.3.1.3	Upper-Level Case and Duct Framing.....	93
4.3.1.4	Condenser Framing	94
4.3.2	Cooling and Ducting Design	95
4.3.3	Electrical and Control Systems	97
4.4	Prototype 1 Evaluation	98
4.4.1	Overview	98
4.4.2	Construction	98
4.4.3	Prototype 1 As-Built Results	102
4.5	Evaluation.....	106
4.5.1	Component Functionality	106
4.5.2	System Mobility	107
4.5.3	Maneuverability.....	108
4.5.4	Survivability	108
4.5.4.1	Ruggedization.....	108
4.5.4.2	Environmental Protection.....	109
4.6	Results	109
4.6.1	System Ruggedization	110
4.6.2	Environmental Protection.....	110
4.6.3	Modular Operation	110
4.6.4	Shore Power Operation.....	111
4.7	Conclusion.....	111
V.	RECON DESIGN AND EVALUATION	112
5.1	Overview	112
5.2	Design Objectives.....	112
5.3	Design Response	113
5.3.1	Framing.....	117
5.3.2	Cooling and Ducting Design	122
5.3.3	Electrical and Control Systems	124
5.4	RECON Build.....	125
5.4.1	Construction	125
5.4.2	As-Built Results.....	127
5.4.2.1	RECON.....	127
5.4.2.2	Computing Stack	129
5.5	RECON Evaluation and Development.....	130
5.5.1	Overview	130
5.5.2	Evaluation.....	131

5.5.2.1	Cooling System Open Circuit.....	131
5.5.2.2	Cooling System Closed Circuit	132
5.5.2.3	Cooling System Development.....	134
5.5.2.4	Computing Benchmark.....	135
5.5.2.5	Computing Benchmark Results	136
5.5.2.6	Electrical and Computing System Development.....	138
5.5.2.7	Mobile Operation.....	139
5.5.2.8	Mobile Operation Results	141
5.5.2.9	Mobility Use-Case.....	141
5.5.2.9.1	Edge Sensor Application	142
5.5.2.9.2	Mobility Use-Case.....	144
5.6	Results and Discussion	148
5.6.1	Cooling System	149
5.6.2	HPC System.....	150
5.6.3	Framing and System Operation	153
5.6.4	Mobility Use-Case.....	153
5.7	Conclusion.....	154
VI.	CONCLUSIONS	156
6.1	Objectives Background	156
6.2	Design Response	156
6.3	RECON Development	157
6.4	Further Research.....	158
	REFERENCES	160

LIST OF TABLES

Table 3.1	80/20 15 Series t-slot extrusion properties.	80
-----------	---	----

LIST OF FIGURES

Figure 2.1	Map of cloud to edge components across the network domain.....	7
Figure 2.2	Decentralized network location from cloud by network hops.....	9
Figure 2.3	Motivation, challenges, and opportunities of edge computing.....	11
Figure 2.4	Energy cost of mobile device transmissions.	14
Figure 2.5	Operational requirements by engineering discipline.....	29
Figure 2.6	Engineering reconnaissance functions.	31
Figure 2.7	USACE mission set.	33
Figure 2.8	Existing ruggedized edge computing resources.	38
Figure 2.9	USACE Azure Stack ruggedized edge devices.	39
Figure 3.1	RECON base attributes and objectives.....	44
Figure 3.2	Concept design progression.....	45
Figure 3.3	NVIDIA DGX-1 front face.	47
Figure 3.4	Diagram breakdown of Nvidia DGX-1 components.....	48
Figure 3.5	Dell Precision 7920 rack workstation front face.	49
Figure 3.6	Internal component view of Dell Precision 7920 rack workstation.	50
Figure 3.7	DDN SFA200NV hard drive unit front view.	51
Figure 3.8	Mellanox InfiniBand SB7700 36 port switch.....	51
Figure 3.9	Pelican rack mount case options.....	54
Figure 3.10	Target Impact Case with standard ruggedization features.	56
Figure 3.11	Impact Target case internal features.....	57

Figure 3.12 Target Impact Case.....	58
Figure 3.13 Impact Target case customized lid features.....	59
Figure 3.14 Impact Target case internal rack features after insulation.....	60
Figure 3.15 RECON ruggedized computing stack.	62
Figure 3.16 Commercial Impact integrated case cooler.	66
Figure 3.17 Portable air conditioning unit.	67
Figure 3.18 Portable cooler with polymer case concept.	68
Figure 3.19 In-row server cooler commercial unit.	69
Figure 3.20 Split cooler design concept with medium duty case and integrated frame.	72
Figure 3.21 Split cooler design concept rear isometric view.....	73
Figure 3.22 Ice Qube split server cooler condenser and evaporator.....	76
Figure 3.23 Ice Qube cooling capacity chart.	77
Figure 3.24 80/20 1515 series extrusion profile and dimensions.	79
Figure 3.25 80/20 1530 series extrusion profile and dimensions.	79
Figure 3.26 RECON integrated component concept design side view.....	82
Figure 3.27 RECON integrated component concept design isometric view.	83
Figure 4.1 Prototype 1 attributes and objectives.....	85
Figure 4.2 Prototype 1 design render isometric view.	86
Figure 4.3 Prototype 1 design render side view.....	87
Figure 4.4 Prototype 1 design render front view.	88
Figure 4.5 Prototype 1 design render top view.	89
Figure 4.6 Prototype 1 framing design render isometric view.....	90
Figure 4.7 Prototype 1 design render framing and hardware bottom view.	91
Figure 4.8 Framing design render side view.....	92
Figure 4.9 Framing design upper level render, top view.	93

Figure 4.10 Case mounting and duct position framing design render, top view.	94
Figure 4.11 Prototype 1 condenser sub-assembly framing design render, side view.	95
Figure 4.12 Prototype 1 ducting and cooling design render, isometric view.	96
Figure 4.13 Prototype 1 cooling design closed circuit airflow.	97
Figure 4.14 Prototype 1 ruggedized computing stack.	99
Figure 4.15 Framing construction Prototype 1.	100
Figure 4.16 Evaporator and partial ducting with framing.....	101
Figure 4.17 Construction framing and partial ducting without side rails.	102
Figure 4.18 RECON Prototype 1, main body computing case, evaporator framing and ducting.	103
Figure 4.19 RECON Prototype 1 side view, evaporator main body.....	103
Figure 4.20 Prototype 1 as-built main body rear view.....	104
Figure 4.21 Prototype 1 condenser system as-built.	105
Figure 4.22 Prototype 1 main body with detached case lid.	106
Figure 5.1 RECON design attributes and objectives.	113
Figure 5.2 Prototype 1 to RECON design progression.....	114
Figure 5.3 RECON design render isometric view.	115
Figure 5.4 RECON design render side view.....	116
Figure 5.5 RECON design render front view.	117
Figure 5.6 RECON framing isometric view.	118
Figure 5.7 RECON framing top view comparison to Prototype 1 framing.	119
Figure 5.8 RECON framing design side view with articulated features.....	120
Figure 5.9 RECON framing design render side isometric view.	121
Figure 5.10 RECON cooling design render with airflow direction.	122
Figure 5.11 RECON ducting panel design render side isometric view.	123

Figure 5.12 RECON design render of cooling component placement.	124
Figure 5.13 RECON design render view of rear control panel access.	125
Figure 5.14 RECON framing during construction.	126
Figure 5.15 RECON duct panel insulation and sealing construction.	127
Figure 5.16 RECON as-built isometric view.	128
Figure 5.17 RECON as-built side view.	128
Figure 5.18 RECON as-built front view.	129
Figure 5.19 RECON updated computing stack.	130
Figure 5.20 Open circuit cooling system test setup.	132
Figure 5.21 Evaporator control unit and refrigerant lines during evaluation.	133
Figure 5.22 Condenser kick-on timer modification.	134
Figure 5.23 RECON benchmark test for cooling and computing systems.	135
Figure 5.24 RECON computing benchmark test setup.	136
Figure 5.25 Benchmark control API for Nvidia DGX-1.	137
Figure 5.26 RECON benchmark test wattage through PDU.	138
Figure 5.27 Mobile operation use-case trailer.	140
Figure 5.28 Mobile operation use-case RECON setup.	140
Figure 5.29 Mobility use-case edge sensor initial configuration with RECON.	143
Figure 5.30 Edge application sensors applied in test configuration.	144
Figure 5.31 Mobility use-case test setup.	145
Figure 5.32 RECON computing stack mounted in trailer during mobility use-case.	148

CHAPTER I

INTRODUCTION

1.1 Background Motivations

The integration of computing, sensors and smart technology across the domain of engineering activities and operations is rapidly expanding. Where a domain is a grouping or visualization of all applications involving a type of activity, engineering activities can be considered under both a physical and network domain. The addition of control systems, devices, sensors and connecting networks provides greater accuracy and information regarding their roles within the operating domain. However, this expansion requires greater processing and physical logistics support to handle high volumes of data, scaling from increased saturation of technology throughout the considered domains. Traditional central computing limits the effectiveness of a widespread integration of technology, because the required bandwidth to transmit large volumes of data across the entire distance between the sensor and computing facility is not consistently attainable. This issue can be resolved by alternate network methodologies which enable onsite data processing and reduce or eliminate transmission requirements. Thus, a decentralized computing paradigm is required to support and expand those operations which fall outside the reach of adequate centralized computing resources.

This project develops a resource within the edge computing paradigm. The term "edge" refers to any conditions, physical and/or network, locations which place the operation outside reach of standard computing resources. Edge computing is a computing methodology wherein

the processing resources are located closer to and in the operating environment, providing low latency processing capabilities while minimizing communication requirements. The general functionality of edge networks, and how they can benefit specialized scenarios better than cloud computing will be explored in the background, leading to the following applications of edge computing.

Edge networks become a necessity for edge of logistics computing scenarios in remote and austere environments which are often seen in military operations. Base facilities, which host engineering operations, exist on the logistics edge, command an ever-increasing quantity of sensors, and demand increased computing performance. Normally, in order to accommodate the processing requirements, the data would need to be transmitted across the globe, subject to bandwidth limitations due to a lack of internet infrastructure or powerful radio transmission facilities. Satellite transmission is time consuming, expensive, and has the potential to be denied due to interference. Additionally, the time it would take to establish a computing facility limits the forward capabilities and development of more sophisticated sensor systems with time-sensitive applications. Unwieldy logistics and delayed payoff can perpetuate doctrine which inhibits the use of advanced information systems. Therefore, with an edge condition applied equally over both physical and network logistics, edge-capable, high performance computing devices are required for the rapid development and implementation of edge networks, enabling advanced sensor systems, and enhancing time-sensitive computing applications.

Under these premises, the research opportunity arises to explore if it is mechanically feasible to create a relatively high-performing computing system capable of operating at the digital and logistics edge using primarily commercially available technology. This system would need to maintain a physical distinction from its operating environment, dust, water, external

temperature ranges and physical forces. The delivery of a computing edge device to the edge location will require design considerations to keep the computing equipment intact, while being transportable and maneuverable during transport and deployment. Could such a system be made largely with commercial components, with existing packaging constraints, cooling technologies and sufficient computing and cooling capacities to perform the required duties in an edge environment? What design configuration can successfully deliver edge computing to edge locations? What are the design objectives to deliver an edge-capable, high-performance computing device which is transportable and maneuverable in various target environments?

1.2 Work Overview

The work contained within this thesis applies to the development and evaluation of a high-performance computing (HPC) node capable of operation in austere environments to fulfill requirements arising from edge computing applications. Feasible concept design is explored using commercial components. Successful design and evaluation of the concept designs and prototypes contribute to the expansion of HPC resources for advanced sensor applications for general and military engineering in edge environments.

Background literature review of this work examined the challenges and benefits of decentralized computing. A comparison of edge and cloud computing methodologies for different scales of computing performance and logistics determined the best approach in delivering expanded processing capabilities to further reaches of geophysical and network location. The selected computing methodology was applied to general and military engineering functions under potential USACE mission sets, determining the design requirements evident for RECON.

The design and evaluation work throughout this effort describes the response to the engineering challenge and requirements of creating RECON. A HPC stack, ruggedized rack mount case, high-capacity split rack cooler and aluminum framing and paneling are selected and integrated into a cohesive unit. Design objectives and priorities in system design are considered and adjusted throughout the evaluation of completed prototypes as they are assessed for providing novel HPC capabilities for its form factor and intended environments.

1.3 Organization

The thesis is arranged accordingly:

- I. Chapter I introduces the conditions of this thesis.
- II. Chapter II is a literature review which examines the background of edge computing relevant to determining the requirements for RECON.
- III. Chapter III begins the concept design narration and process of responding to the requirements, establishing design objectives and priorities.
- IV. Chapter IV outlines the design details and evaluation of the first RECON prototype.
- V. Chapter V responds to the evaluation of prototype 1, outlining the design of RECON, and the full evaluation of RECON across various tests and use cases.
- VI. Chapter VI concludes the background, design and evaluation results of this thesis.

CHAPTER II

BACKGROUND

2.1 Edge and Decentralized Networks

Edge computing network architectures present a great opportunity to offer different operating characteristics to the traditional centralized cloud-based architecture for data and processing access. Edge computing architectures can create and optimize network functionality where cloud architectures are lacking. The advantages of creating a high-performance computing node at the edge of networks can be understood through the comparison of cloud to edge and fog computing concepts. Additionally, the discovery of advantages and disadvantages in the general concepts of public decentralized networks can inform the specific motivations and opportunities in private or specific applications, such as military operations support.

Edge computing shifts the concentration of services and computing to the outer reaches of the network, across network and physical operating domains, decentralizing computing and communications from standard central cloud nodes. In relation to cloud structures, where the data centers, servers, databases, and storage are the main access points which all data and communication are routed through, up and back down the chain of devices, both users and devices may instead see data routes and processing done in non-centralized network locations physically closer than the cloud locations. The functionality of edge and related decentralized networks are differentiated by the manner of processing and communication done at the edge nodes. Within the Internet of Things (IoT), referring to the

devices which interact with reality that can connect with each other and the network, edge refers to the landscape of the local network where sensors and IoT devices are located [1]. The edge is the first hop from the IoT devices such as the Wi-Fi access points or gateways [1]. Further specifying the decentralized network methodologies, if IoT devices interconnectedly perform the computations themselves, without involving any organizing network device, it is referred to as mist computing, whereas edge computing utilizes all available components within the decentralized network domain. This horizontal communication among edge proximity devices, compared to vertical communication can be visualized in the figure separating the network of devices both by physicality and network centrality.

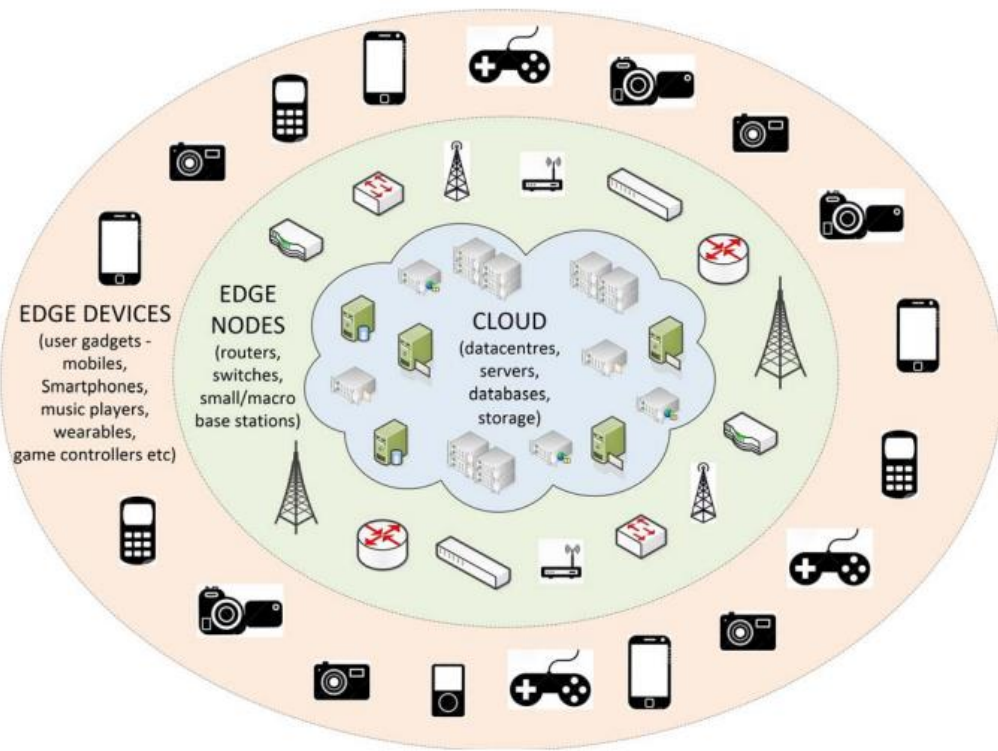


Figure 2.1 Map of cloud to edge components across the network domain.

Edge devices to centralized cloud resources visualized in “Challenges and Opportunities in Edge Computing” by Varghese, B. et. al., [2]. Typical communication of cloud services travels linearly to center, while edge computing allows horizontal or radial communication across edge devices and nodes.

The network domain is populated by the different device structures, varying by functionality, purpose and physical location and interaction with end users and effects. Central processing structures are the cloud datacenters, servers, databases and storage as shown above. Edge computing begins outward over network and geophysical location, in the green and orange fields shown on Figure 2.1, where edge nodes connect cloud structures to the edge devices. The connecting edge nodes, as shown, are routers, switches and sub stations, interfacing between the individualized devices, the edge devices such as mobile phones, smart devices, wearable devices,

sensor equipment, and the robotics platforms which are essentially anything on the edge of the communication line of the network and the physical edge of the information domain. Typically, the centralized communication structure would flow outward and back through the layers, from the most outwardly point on the edge all the way to the center, or in Figure 2.1, about the radius as opposed to the angular location. To change the direction of data flow and processing to move in an angular fashion exclusive to the outer layers of the network domain, between edge devices and edge nodes, is the idea of operating on an edge network, where the aim of such network designs is to perform computations on nodes through which the network traffic is directed through routers, switches, and base stations, referred to as edge nodes [3]. A linear comparison of the various computing paradigms is shown below, where the location and range of the computing paradigm devices are applied across the range of network hops, from edge to cloud.

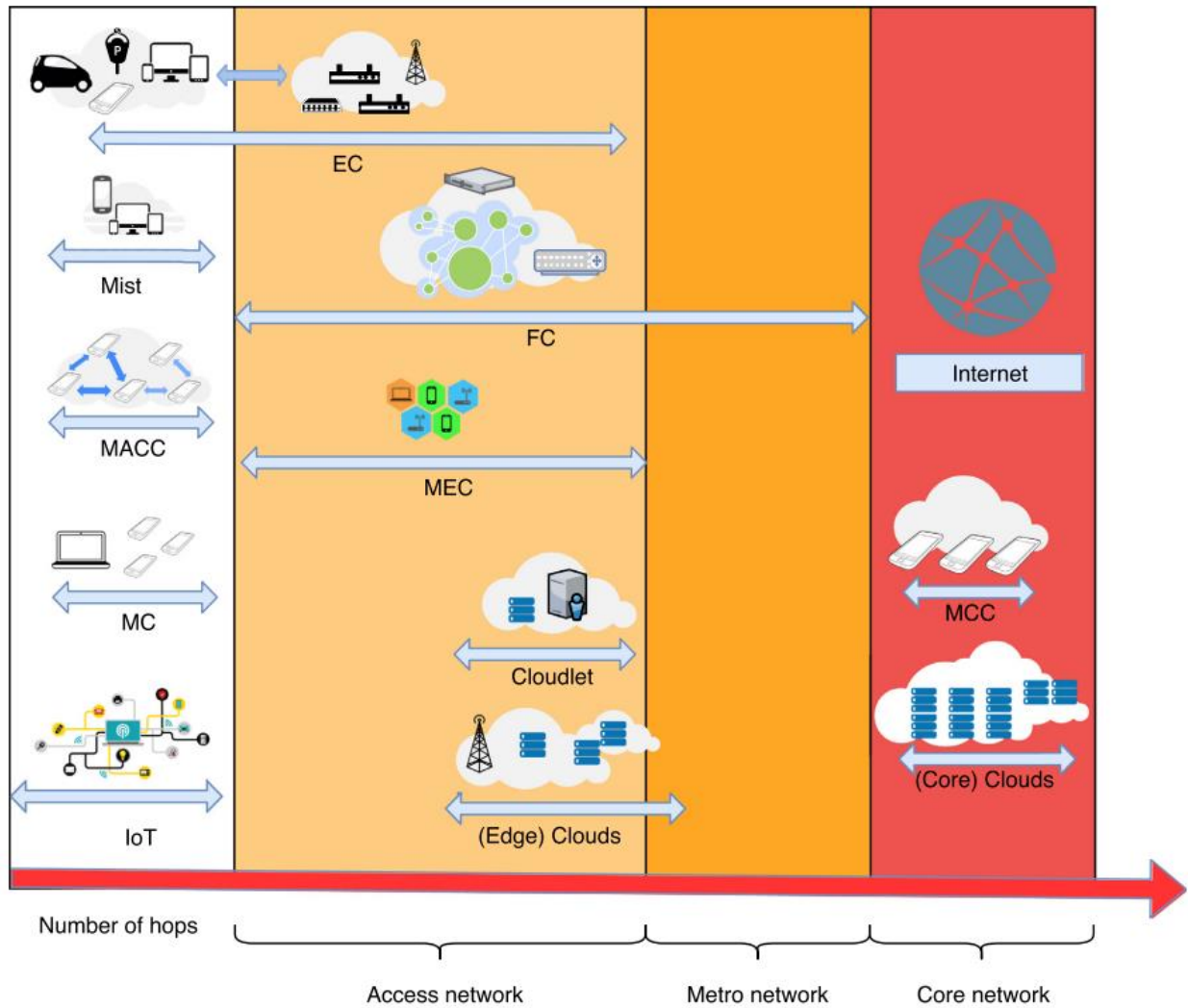


Figure 2.2 Decentralized network location from cloud by network hops.

A comparison of computing paradigms in terms of their specific network location and distance from the centralized cloud network. Their location is also applied over the general network hops required to communicate across the network domain, visualized in “All one needs to know about fog computing and related edge computing paradigms: A complete survey,” by Yousefpour, A. et al., [1].

Edge computing takes the domain within the edge devices at IoT level, and outside the cloud computing center, covering each step between the central and outermost locations. The Edge Computing (EC) shown above is the network between IoT devices and the connecting devices throughout the first hop above. The edge network different than independent network

operation of the IoT devices and mobile devices. Networks formed exclusively by those devices, with shared computing done on IoT devices themselves is referred to as Mist computing (MC) [1]. As a connecting structure, edge computing is a crucial computing paradigm in the current landscape of IoT devices because it integrates the IoT devices with the cloud by filtering, preprocessing, and aggregating IoT data intelligently via cloud services deployed close to IoT devices [4]. The manner in which edge nodes are configured to process and communicate data can be leveraged according to what attributes of network performance need to be optimized. Thus, edge computing methodology provides the most flexibility among decentralized paradigms for the improvement of network performance.

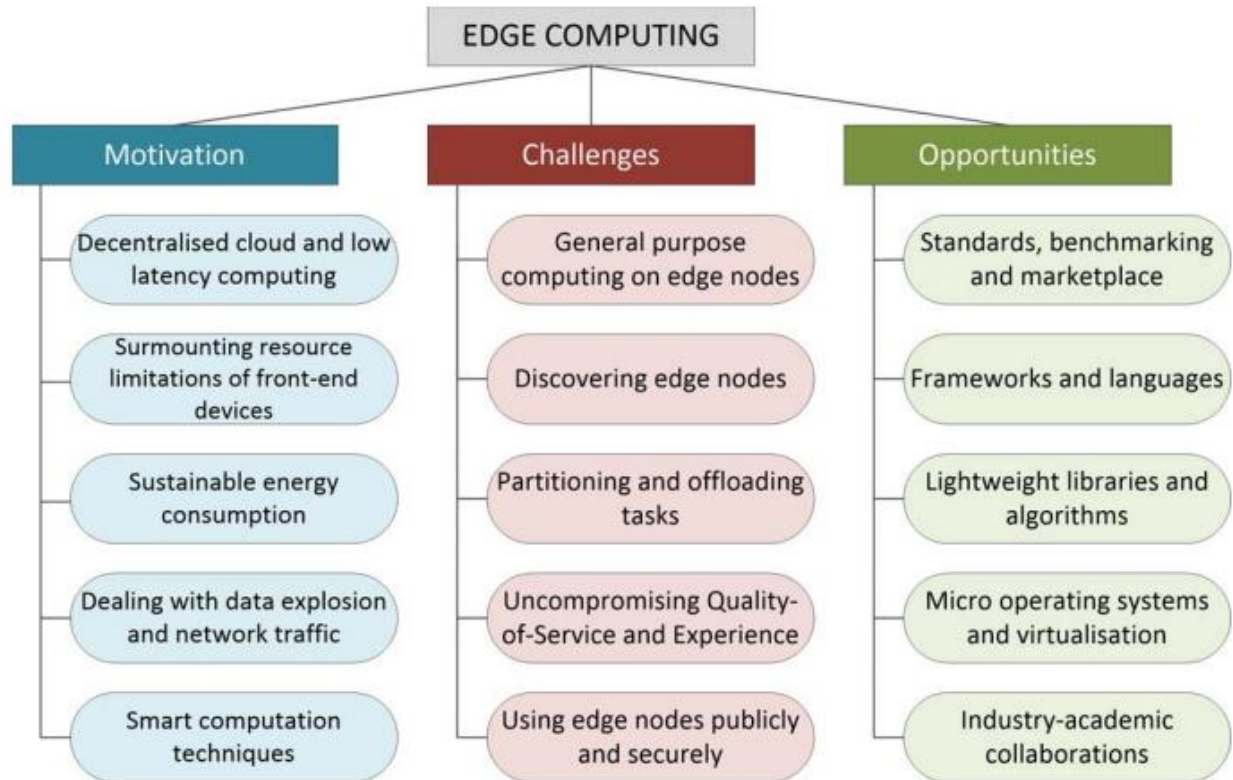


Figure 2.3 Motivation, challenges, and opportunities of edge computing.

Overall motivations, challenges and opportunities of edge computing networks identified in “Challenges and Opportunities in Edge Computing” by Varghese, B. et al., [2]. The general benefits and associated challenges of edge computing can be compared in order to determine scenarios in which edge computing is advantageous over cloud computing.

Edge computing applied over general networks provides a large set of changes to the way a network can perform. As illustrated in Figure 2.3, the immediate benefiting functions of edge networks are motivations for its use. Long term operation and development of edge networks can be directed towards identified opportunities, further enhancing the immediate benefits. The best way to ensure successful implementation of edge networks in the general network domain is to capitalize on the benefits and address the challenges. Benefits for the general network include the reduction of latency from cloud resources, reduction in bandwidth requirements, a more flexible network traffic capability, and adaptable processing techniques, among other benefits.

2.2 Edge Computing Motivations and Benefits

2.2.1 Latency

The primary motivation and benefit of edge networks is in latency reduction. The reduction is achieved by shortening both the distance and volume of data transmission over network logistics. Central cloud locations require greater transmission distance and network device for each communication, from the edge devices, through the edge nodes facilities until the cloud facility. Additionally, each single edge device must complete the full transmission cycle across that distance to complete each communication. Groups of edge devices in location will all individually stack up communication bandwidth across the entire line of communication to the central facility, requiring more overall network resources to support. The required bandwidth must be held up for the total of edge devices, requiring ever greater expansion of the network logistics capability for maximum bandwidth through the entire line, otherwise resulting in bottlenecks of service, and degrading the end user quality of service. This issue is critical for applications requiring low latency at geographically far locations from the cloud service. Particularly latency sensitive applications converging multiple streams of data for real time analysis, such as augmented reality glasses or visual guiding service using a wearable camera require response times between 25ms to 50 ms. In VR and other applications which interact with the senses, particularly low latency is required. Humans are acutely sensitive to delays in the critical path of interaction, as human performance on cognitive tasks is fast and accurate [6]. For example, under normal lighting conditions, facial recognition takes 370 – 620 ms depending on familiarity, and speech recognition takes 300 – 450 ms for short phrases, only requiring 4 ms to recognize that a sound is a human voice. In VR applications using head tracking, latencies are required to be under 16 ms to maintain perceptual stability [5]. Such speeds and requirements

highlight the limitations of cloud services. Averaged examples of latency studies of cloud computing services, for example, by Ang Li and colleagues, reported that the average round trip time from 260 global vantage points to their optimal Amazon Elastic Compute Cloud (EC2) instances is 74 ms [7], which would be in addition to the latency between the edge device and network connection. At the extremity of distance communication, Berkeley, California to Canberra, Australia, the latency is approximately 175ms. Thus, cloud infrastructures present a challenge to the latency requirements of edge devices using cloud services, making real time decision making not possible when depending on processing of live data streams and sensor input [8] [10].

2.2.2 Communications Transmission Cost

Edge networks address the latency challenges by decentralizing the required communication distance between the edge device and processing location. In addition to latency caused by network distance, computing and processing requirements can be reduced by leveraging edge nodes to prepare and filter the data processing at central cloud services. Edge computing nodes reduce the hops and network facilities that the communication must travel through geographically and in network. Should infrastructure be deployed to manage the interaction of edge devices with each other and with cloud resources, the data processing latency seen by cross world communications would shrink to low latency levels acceptable for human perception in VR applications and sensor interaction. The cost for communication response time and logistics can be reduced, shown in Figure 2.4, by comparing response time and energy costs per operation of a facial recognition and augmented reality application, on mobile devices to Amazon Web Services datacenters. Cloudlet processing, a cloud service geographically closer

than main cloud facilities although not decentralized enough to be part of edge networks, reduces processing operation time and transmission energy cost.

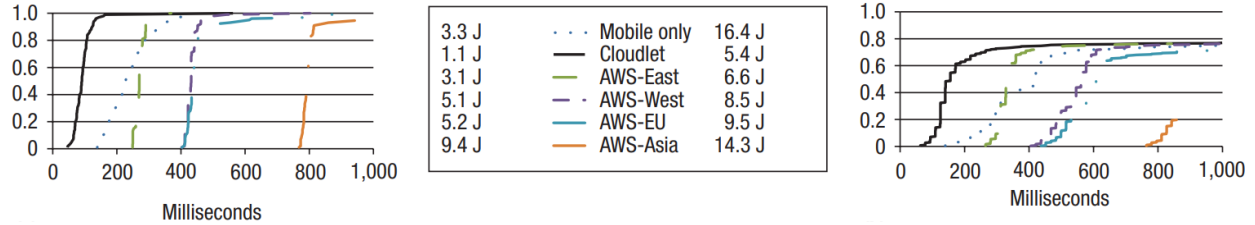


Figure 2.4 Energy cost of mobile device transmissions.

Energy cost of cross hemisphere transmission indicate the importance of offloading data processing from mobile devices and cloud service can increase processing efficiency, reviewed in “The Impact of Mobile Multimedia Applications on Data Center Consolidation,” by K. Ha et al., [10]. Response time distribution of an augmented reality application on the left, and a face recognition application on a mobile device on the right, along with the respective energy costs per transmission for both, center, are shown. Transmissions are sent over a WiFi first hop to the cloudlet or an Amazon Web Service data center. Closest to edge node functionality, cloudlet displays best energy and latency performance [10].

The collective costs of passing communications through the necessary network logistics to the centralized operating center is higher than edge network costs. If an edge network can handle the local operations, a movement toward edge infrastructure will increase the quality of service for those applications. Configurations which leverage existing edge nodes and add accessible computing resources to edge devices will reduce the amount of full cycle transmissions to cloud facilities and can otherwise simplify and reduce the data and bandwidth of transmissions to central facilities.

2.2.3 Smart Data Management

The management of data across multiple resources, on an edge device, node or cloud is possible with smart techniques. Data management configurations leveraged across the devices with the available computing power in the edge network compensates for the limited power of edge devices, such as mobile phones and battery powered devices. Edge devices, over the network domain of operations, connect the network to the physical domain, operating as sensors, feeding data converted from physical phenomena. Sensor devices on the edge, such as mobile devices, are able to capture all the various data inputs such as text, audio, visual, pressure, motion, temperature etc. These front-end systems and devices either save or send the data further into the network. However, these front-end devices cannot perform complex analytics due to middleware and hardware limitations [11]. Similarly, the data captured by compound sensor systems cannot be processed on the front-end devices themselves, requiring cloud services for complex data analysis. The complex analysis requirements can instead be filled by edge computing nodes on network, or edge computing devices tied close into the local network. Consequently, analytical workload could then be processed with edge network, making complex sensor data useful in context to local formations of sensors.

2.2.4 Edge Device Processing and Energy Management

Offloading complex data processing from front end devices and other edge devices saves battery and energy use when complex computations can be run in more efficient configurations. While batteries afford useful mobility, battery life limits the processing effectiveness of edge devices compared to established energy logistics of edge node facilities and larger processing devices. Direct cloud center processing is also efficient; however, local edge nodes provide greater battery and energy efficiencies when the transmission of data to cloud facilities requires

significant signal strength. For example, edge devices which are in mountainous or signal blocked urban areas must amplify their wireless signal output in attempt to communicate with those wireless stations, thus draining battery. They may otherwise need to rely on satellite communication for their cloud data processing, again requiring larger energy use per transmission, and draining battery resources. In instances with an established need for data processing in communication denied areas, local edge networks can reduce the need to communicate with cloud resources and alleviate data processing on battery devices. Smart distribution of data workloads along the processing pipeline, such as first hop processing done on local routers and local edge computing nodes importantly minimizes processing done on battery power. Generally, edge nodes can facilitate computations nearer to the source of data, or where data is generated, and can incorporate strategies for remotely enhancing the capabilities of front-end devices [2].

2.3 Edge Computing Challenges

Given the benefits and motivations for shifting the IoT towards an edge computing paradigm, challenges arise in the implementation of the paradigm shift. Existing advantages of cloud facilities remain in contrast to the challenges of edge computing when edge network logistics are introduced. The context of each situation determines if the benefits outweigh the particular challenges associated with edge computing applications. Edge computing must then be deployed within the regions of geophysical and network domains where the benefits outweigh the challenges.

2.3.1 Device Inflexibility

The emergence of edge computing among the predominantly cloud computing landscape presents challenges in establishing the base functionality that it seeks to replace. General computing, edge node discovery, task partitioning or smart computing, edge logistics, dependability and security are the challenges highlighted by establishing edge computing environments. Using existing network nodes of the cloud communication infrastructure may subject them to general computing loads for which they are not suited. The network infrastructure between edge devices and the cloud, such as access points, base stations, gateways, traffic aggregation points, routers, switches, etc. are optimized for existing cloud functions. Base stations use digital signal processors (DSPs) that are customized for the workloads they handle, and consequently, are not suitable for generalized workloads because of their specific design [2]. This specialized infrastructure is a constraint in edge computing due to their potential inflexibility. Therefore, edge networks need to be configured to take advantage of the aggregate groups of different computing resources when there is an opportunity for it, outside of cloud communication priority. Developing agile protocols for the conversion of existing cloud-edge resources is required for a widespread enhancement of edge device functionality.

2.3.2 Appropriate Processing Matching

Established decentralized edge networks run into the next challenge. With multipurpose processors from edge node processors forming a wide network of connection points across the network domain, it is not clear at which point the edge devices computational load should be processed. While the discovery of resources and services in a distributed environment is well explored, the sheer volume of the decentralized edge network exceeds the simple manual network and service detection mechanisms seen in typical cloud network discovery [2].

Additionally, a wider variety of computational loads, such as neural network processing or large-scale machine learning will complicate the processing requirements, thus increasing the challenge to rapidly discover appropriate nodes at which to process the new computations. An effective edge network paradigm must deliver on the benefits of low latency and more efficient data processing for edge devices, upholding seamless integration and removal of computational nodes for the user experience. Achieving rapid and robust edge network navigation and discovery of processing nodes unlocks benefits offered by decentralized processing.

2.3.3 Edge Node Seamless Discoverability, Scalability

The cumulative challenges of connecting edge networks affect the cumulative condition of the user experience. Upholding an acceptable user experience requires resolving multiple challenges of edge computing to maintain a seamless experience. Considering that the purpose of edge computing is to improve network function and operation at user discretion, the improved applications supported by edge computing must be dependable and constant by the user's perception. New applications, including augmented reality (AR) with two-way communication, require low latency. Low latency performance from edge computing is necessary in AR applications, where latency needs to remain below 16 ms for 60 frames per second during user interactions [12]. With the expectations set, low latency AR applications on the edge computing network require seamless operation. The challenge of maintaining acceptable user experience pulls in issues of edge network scalability, and the earlier mentioned aspects of task distribution and node discovery become relevant. Scalability of service for edge nodes is a concern, where the edge infrastructure does not flexibly increase in processing power when the local users increase suddenly, for the expected service. While latency cannot be directly improved under cloud services, processing power is immediately scalable to demand however, the

communication bandwidth needs to be sufficient for data transmission. Additional edge nodes can be pulled into the processing pipeline, but they must be discovered and configured for the scenario's application. Edge computing networks will be challenged to provide seamless network service to applications pushing the limits of local computing resources when those applications also require top performance attributes from the edge network.

2.3.4 Seamless Operation Over Geospatial and Network Boundaries

In addition to the issue of scalability over local network capacity, the continued smooth operation of edge applications over the network will be challenged to continue that operation when the location of the operation changes. Changes in geographical locations correspond with changes in the network location. Similar to the challenge of task partitioning and node discovery, the locations of edge facilities typically do not change, and the processing pipeline for edge applications must be rapidly adjusted. The change in pipeline is more significant than the routing for typical cloud network connections, which is a well-established process in mobile edge devices. Seamless operation is an additional challenge in edge computing, when the edge device traverses outside the bounds of any single or particular local group of the edge network configuration. Edge computing devices, nodes, gateway, etc. in the network and physical domain are connected by discrete and exact bridges between the two devices. Effective movement in the network domain requires connections to be broken and reformed into new device configurations and pathways, thus creating interruptions, leading to seams in service. Movement in the physical domain is more flexible with the use of wireless data connections, local area networks and wide area networks. While the connections made by physical phenomena are maintained, the network domain connection remains uninterrupted. Edge devices equipped with the capability to connect to network devices, with a physical aspect of mobility within range of the established edge node

network connection chain can better operate in the environment seamlessly. Edge devices used in AR applications can traverse the physical domain to an extent with the users, while maintaining the current network position, giving them an aspect of mobility without interruption. If the physical mobility of edge devices can be used in conjunction with mobile edge network components and nodes, then the challenge of seamless service to intensive edge applications can be addressed. Successive physical mobility of components along the network path, starting from the edge devices towards central network, effectively increases the operating range of edge devices and applications with seamless service. If the physical location of the branch of components along the edge network can be adjusted during operation to accommodate the edge application, the effective seamless service capabilities can be extended for specific use cases.

2.3.5 Edge Component Location

The flexibility of edge networks across geographical locations also opens the network up to additional challenges inherent to their location. Edge network devices will be deployed in various locations across the operating geography. They may be placed in public or private locations, and each node may be of greater or lesser operational importance in service width and accessibility. Proportionally more resources must be spent in order to access and maintain the devices spread across logistical space, compared to the cloud facilities. The decentralized nature of edge devices makes them difficult to secure from intentional threats, in addition to natural wear and entropy. In a nod to the anticipated large scale of these systems, the security mechanisms themselves must be scalable and decentralized [12]. Centralized cloud resources are concentrated in main facilities, where they are easily maintained, accessed by authorized forces, and are additionally secured against digital and physical threats. Edge nodes and components are not secured by those centralized means, and maintenance of those devices incurs additional cost

when devices need to be tended to at each location. Component failure at edge locations add to the challenge where components cannot be hot swapped like they would in centralized facilities. As hot swap components are not practical in edge environments, edge nodes such as repurposed routers in home environments, may alleviate the failure points by forming aggregate networks with other routers, negotiating peer-based failover in the community [12]. This alternative will still require robust data security protocols and smart task partitioning, providing additional challenge to the edge scenario. The deployment of edge devices as network infrastructure would multiply cost based on individual unit hardening requirements. While edge node devices would likely be cheaper devices more suited for mass distribution, standards of physical security and hardening against physical phenomena are required to gain a standard quality of service to ensure important operations are trusted through the network. Hence, Edge devices and nodes must have additional provisions for the necessary hardening and protection and maintenance required for operation in edge environments, with their survivability and resiliency scalable to the intensity of the edge application.

2.3.6 Security

The decentralized edge network invites security issues over the network domain similar to the security issues of the physical domain of operation. Discussed in [12], decentralized management and device tampering can lead to chain of trust and security or privacy violations. This possibility must be addressed in both small- and large-scale edge computing environments. Cyber security solutions can be extended from cloud computing to edge computing environments. Cloud techniques of authentication, access control and distributed intrusion detection need to be adapted to the edge environment to address the security challenge. Attackers can more easily tamper with edge network components because they are distributed closer to

clients and cannot be protected, affecting trustworthiness and authentication. Established public key infrastructure (PKI) is computationally expensive to apply across smaller edge network devices. Management of the large number of keys generated by an IoT environment also add to that challenge. Additional layers of security, applied at device level, such as biometrics and log in expiration timings would reduce risk. Edge device ecosystems would need decentralized security measures to be consistent, as the nature of edge networks spread over geophysical locations can result in intermittent connection to central security processes. Peer to peer security systems and decentralized security tailored to the edge devices as needed per use case will ensure edge computing can achieve some level of autonomy in security. The challenge of security will need to be considered over a standard of overall risk allowance, or in specific configurations applicable to the objectives of a local edge network group.

2.4 Edge Motivation and Challenges Conclusion

The general motivations of edge computing are counterbalanced by their challenges. As edge computing is in infancy compared to cloud computing, the introduction of edge infrastructure comes with additional growing pains natural to creating new systems. Specific edge computing groups deployed to target environments, where the challenges are accounted for, allow for provisions to address concerns of hardening, security and network functionality. Such edge computing groups developed for specific operations and objectives, where the benefits to challenge balance is favorable, spearhead the application of edge computing into the physical and network domain.

2.5 Edge Requirements for Computing

New deployments of systems using edge computing require new computing equipment, hardened and purpose built, to be deployed on site of the application without existing facilities. Alternatively, they must leverage existing communications infrastructure, converting the functionality of cloud infrastructure into a broad edge computing network. These different scales of network conversion require different costs to setup, the first option being the costs of setting up a deployment of new equipment to the scale required of the edge application, and the second, a broad conversion of systems and protocols to enable edge network functionality in addition to the existing requirements of cloud communication. This comparison of edge network deployments in new environments to the edge computing paradigm is a difference of two scales of edge deployments: One leverages the concepts of a large-scale existing infrastructure conversion, and the other, a specific computing group purpose built to the application. These scenarios provide a comparison for either success or defeat of edge computing in new environments.

2.5.1 Cloud Infrastructure Edge Conversion Configuration

. Existing cloud infrastructure can be converted into edge network support for local infrastructure and applications. In scenarios for the introduction of edge computing in cloud centric computing infrastructures, existing computing resources can be reconfigured to support edge computing. This approach can leverage network infrastructure in a small community-scale up to industrial-scale cloud infrastructure physically located nearby. Community-scale, for example, could use the local routers in a microgrid for their spare computing power. The latency benefits are gained, and existing resources can be leveraged without large investment in new infrastructure. Establishing the use of local routers to enhance edge devices in a community scale

can support edge applications such as augmented reality with the low latency and reduced device power consumption benefits, but now without challenges. The reconfiguration, security and resiliency, as well as processing capability are all challenges with this approach. Configuring existing infrastructure can be as difficult as installing new equipment to establish the edge network. Edge nodes sourced in community routers and small devices exist in varying conditions. If access to those devices is allowed, the dependability of the devices would not be held to any high-level standard of operation. Security and reliability risks are pronounced for existing infrastructure conversions, and the application of these edge networks would not provide high security and dependability standards. While converted cloud infrastructure provides lower quality computing resources, the introduction this edge network configuration can add broadly available edge resources, assisting in the growth of edge computing.

2.5.2 Small-Scale Edge Computing Configuration

Edge applications which require high quality of service, dependability and security must employ specifically configured edge node systems for the scenario. Instead of leveraging existing network infrastructure, new network and processing equipment can be installed for best effect in demanding edge scenarios. Edge computing scenarios such as the consolidation and processing of high amounts of data from edge devices for AI classification and sensor integration, AR communication outside of established network domain infrastructure, or physical domain infrastructure for that matter, require more focused and intensive computing support than what cloud infrastructure can offer. Task specific edge network systems over a small scale should be deployed when considering scenarios where the edge application requires edge network benefits such as seamless quality service and low latency processing, while addressing the challenges of heightened security and survivability, with less importance placed on smart

computational techniques and flexible task offloading over the edge network. Small scale edge deployments best cover edge network scenarios which also operate on the edge of the network domain as it exists within the physical domain, meaning that the scenario already has limited or no access to existing network infrastructure, and can place the computing equipment in areas with additional physical challenges against computing equipment. Discussed in the challenge of addressing seamless quality of service, task specific edge computing configurations can be tailored towards different issues in the logistics of the computing components. For example, scenario requirements including computing component mobility, survivability and communication security over its operating area can be addressed. Extended seamless service range can be provided by introducing an aspect of mobility to hardened edge node components. The edge device and first hop edge node specifically can be mobilized in order to extend seamless service range for local edge device groups addressing geographical concerns. In those ways, task specific small scale edge network setups can introduce high power edge networks into areas without existing network infrastructure, coincidentally solving logistical issues of network communications and capabilities, furthering the edge of the network domain over physical boundaries.

2.5.3 Edge Computing Configuration Comparison

Overall, edge computing configurations perform differently by scale. Large scale conversion of existing cloud infrastructure can glean processing support for edge devices and lighter applications, however, lack processing power and dependability due to the varied purpose and condition of existing infrastructure. When comparing the overall challenges to benefits of edge computing, large scale edge computing which leverages existing infrastructure is advantageous when the challenges of security, reliability and seamless service are not critical

factors for lighter weight processing applications. The cost of being able to repurpose various infrastructure equipment is mitigated by the introduction of flexible edge networks to large areas of the network domain, where the edge computing paradigm can grow, proving itself flexible to multiple computing purposes. This is contrasting to one-way directional communication to cloud resources, as edge devices, routers, etc. can communicate directly with each other, greatly opening the directions of communication in that area. Small-scale edge deployments become the best option in edge application where high processing power, dependability and seamless quality of service are required, although under a more limited range, even in areas without any existing network infrastructure. Challenges of applying general purpose computing, node discovery, offloading and partitioning of tasks are eliminated when the edge computing group is designed for the target environment. These scenarios demand low latency, high performance computing with seamless service and security. Additionally, the edge equipment must be deployed due to the inability of existing network infrastructure, meaning that the system can bring an intensity of edge computing to new areas at force, despite the cost it takes to harden and power the setup as needed. Small scale edge networks are less flexible in connecting to different edge nodes under the considered edge scenario and do not add as much to the network domain, however, they can add network connections in areas previously unreachable.

2.5.4 Edge Environment Requirements

Edge computing provides many benefits across computing needs, while also offering challenges in its implementation. In a cloud dominated style of networking, edge computing is an emerging technology, and needs to first establish itself in new computing environments, either digitally or physically, as explored in the small to large scale styles of edge computing deployments. Edge computing deployment can be better applied and understood as a problem of

logistics. Logistics, generally applicable in practice and industry, is also applicable to edge computing, as edge computing is a methodology to find a way to deliver the best processing power and quality to the end user with the lowest latency. In doing so, edge computing must solve the task of how to deliver the computing power in the network and physical domain, which introduces the edge paradigm into the new domain locations. To enhance the performance of the furthest edge devices, edge infrastructure must be logistically viable in delivering edge computing to the furthest scenarios in order to expand the network domain's edge of operation. To operate in the new edge environments, edge nodes must be hardened and configured to support new edge applications at the edge of computing.

2.5.5 Edge Configuration and Requirements Conclusion

The broad opportunities of edge computing are applied to applications and scenarios, bringing the edge equipment into the operating environments, and subjecting them to operational constraints of the environment. The constraints define the required capabilities of edge equipment, ultimately defining the required attributes for successful deployment in said location, scenario and environment. The logistical requirements of getting the edge equipment to the operating location is layered on top of the existing operating requirements, where equipment must be able to interact with existing logistics capabilities. Logistics requirements define remaining aspects of capabilities that an edge node must address. Edge scenarios and applications in austere environments additionally define the requirements of small-scale edge deployments, for example, as those environments likely do not have the physical infrastructure to protect equipment from exposure. Logistics requirements for the edge computing resource's timely and efficient use add to the requirements. Specified geophysical location, operating

scenario and associated logistics constraints of the edge computing application define the design requirements, connecting the edge computing opportunities to reality.

2.6 Engineering Mission Objectives Background

Small-scale edge computing setups generate opportunity to apply edge computing in various environments and mission spaces. The scenarios and environments that the edge computing systems are meant to operate in should be defined. Given the connection between the opportunities of edge computing to requirements of small-scale edge computing system deployments, with additional requirements imposed by geophysical location and logistics capabilities, the surrounding context and objectives define edge systems. The mission space of the USACE is the primary objective for RECON and other edge systems examined.

The broad mission set of USACE can be used to identify opportunities in edge computing, sequentially revealing the application, environment and requirements of small-scale edge computing deployments to edge environments. The USACE mission is to deliver vital engineering solutions, in collaboration with our partners, to secure our Nation, energize our economy, and reduce risk from disaster. The mission to support those challenges will benefit from enhanced edge computing applications. Edge environments, the location in both the network and physical domain where adequate access to traditional computing resources is limited, considerations for logistical complications must also shape the requirements. Limited aspects of engineering operation requirements from combat engineering, general engineering, and geospatial engineering, as well as operations of facilities management, civil works and emergency response give the scenarios in which opportunities of edge computing can be applied to provide support.

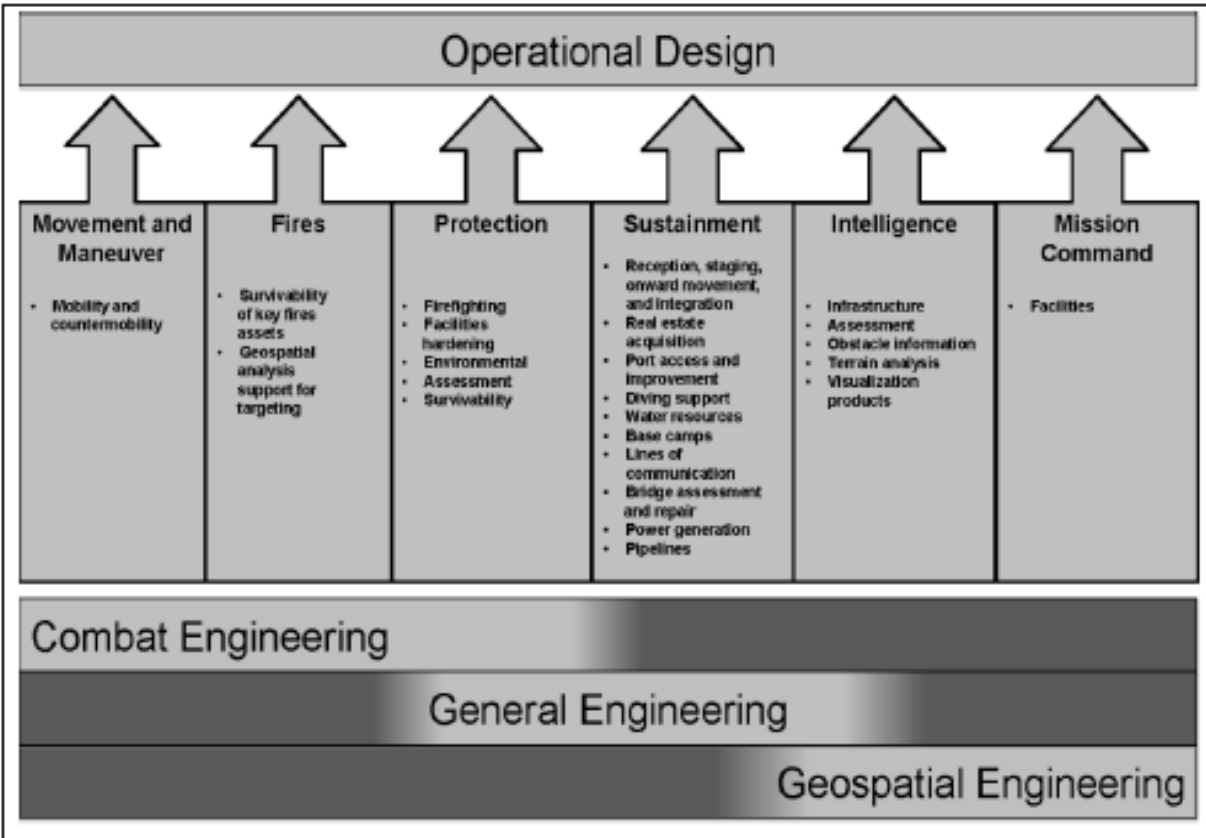


Figure 2.5 Operational requirements by engineering discipline.

Operational areas over general and geospatial engineering can be considered for edge computing opportunities.

Broad theater engineer requirements are spread across combat engineering, general engineering and geospatial engineering. These functions cover operation types shown above. These are movement and maneuver operations, fires, protection, sustainment, intelligence and mission command, each with objectives. Applicable objectives under each function can be supported by the application of edge computing, supporting the application of sensor technology and the coordination of technological resources. From left to right, the immediately applicable objectives are supporting mobility and counter mobility, ensuring survivability of key fires assets, geospatial analysis support for targeting, supporting firefighting, facilities assessment for

protection, hardening, survivability, and environment assessment. Edge computing can be applied to sustainment objective, in the assessment and monitoring of sites for reception, staging, onward movement and integration, port access and improvement, driving support, water resources, base camps, bridge assessment and repair and power generation. Intelligence operations will benefit from edge computing support for infrastructure monitoring, assessment, obstacle information, terrain analysis and visualization of data. Finally, facilities monitoring can be supported by edge computing deployments. There exists a large variety of operations in which edge computing can support the technology that enhances the execution of those operations, in either military or civil applications.

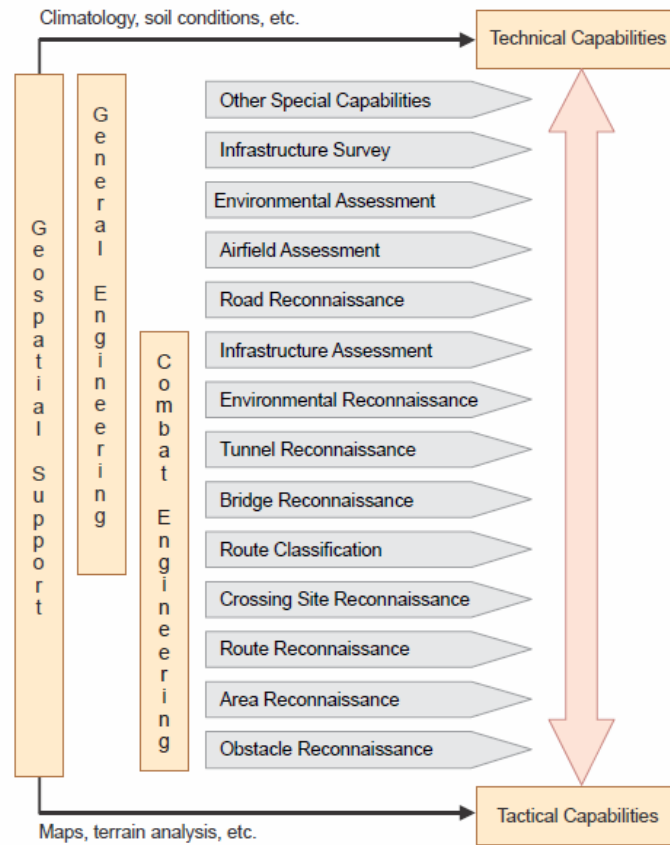


Figure 2.6 Engineering reconnaissance functions.

Engineering Reconnaissance and engineer functions that can be situationally supported by edge computing deployments.

Additional example of mission space and applications are shown in Figure 2.6. A key edge application is the aggregation and processing of sensor data in edge environments.

Reconnaissance and engineer functions and their current solutions can be enhanced by small deployments of edge computing equipment for rapid development of sensor data. The various engineering reconnaissance and assessment functions above can be enhanced by edge computing, given the ability for an edge node to be deployed to the environment within a useful time frame.

2.7 USACE Mission Set Background

The broad base of activities and operation conducted by USACE calls for a wide variety of operating situations and environments, which can be simplified into either disaster or military environments, establishing the expected operating environments, in order to define requirements for RECON Emerging technology and the general reduction of sensor prices allows for the greater use of these technologies in all areas of operation, increasing the capability to collect large amounts of data in the field. Field conditions are varied, when applied to support military missions, civil works in varying continental weather and disaster logistics conditions, and in computational intensity of geospatial support and research and development. Examples of scenarios in each mission space, covering engineering operations and reconnaissance as well, will provide a range of environmental and logistics factors to consider as edge environments.



Figure 2.7 USACE mission set.

The broad base of activities and operation conducted by USACE calls for a wide variety of operating situations and environments. Emerging technology and the general reduction of sensor prices allows for the greater use of these technologies in all areas of operation, increasing the capability to collect large amounts of data in the field. Field conditions are varied, when applied to support military missions and civil works. All of which are conducted in varying inclement weather and disaster logistics conditions, and in computational intensity of geospatial support, research and development. Examples of scenarios in each mission space, covering engineering operations and reconnaissance as well, will provide a range of environmental and logistics factors to consider as edge environments.

2.7.1 Military Environment Edge Computing

One of the primary edge environments, including military missions in construction, installation support, reconnaissance and object assessment, as well as general operations engineering, is arid desert environments. Essentially the desert theater, many operations are conducted in dry, dusty, arid environments with lower levels of logistics infrastructure and support. Technology supporting operations in those environments require hardening against the environmental conditions and hazards introduced by logistics and operations, excluding direct combat related damage. High temperature variations can cause fatigue across different materials. Dust and sand ingress is a constant risk for components not sealed against it. Transportation across terrain can be jarring and damaging to shock sensitive components, and existing road infrastructure can add to those concerns. Edge computing devices need to consider those environmental factors.

Engineer operations outside of the contiguous United States (OCONUS) in arid environments include the various aspects previously outlined require HPC resources in GPU intensive applications Reconnaissance, assessment, security and analysis of facilities and infrastructure are already enhanced by sensor technology, where various visual, thermal, vibration, and location data is integrated into reconnaissance functions. Enhanced reconnaissance functions provide sensor feedback to better communicate accurate data of the environment, utilizing high fidelity aerial and ground-based photogrammetry to capture 3D terrain and environmental data. Processing that data through standard computing workstations in order to stitch together various data types, as well as to run artificial intelligence object classification, may take days if not weeks to process with standard PC resources typically in edge environments. Additionally, enhanced sensor capture of terrain and environmental data can be

used for virtual reality training and physics model visual applications, all of which would benefit from HPC resources for quick support and data processing. Flexible HPC and GPU computing resources can reduce the timeframe of processing by a factor of 20, allowing for high fidelity terrain data to aid in decision making when terrain models can be constructed promptly.

2.7.2 Disaster Environment Edge Computing

Missions responding to disaster scenarios and support across the contiguous United States will expose edge component and applications to more varied environments. Disaster response will subject edge devices and components to wet and humid environments following storms. Edge devices in those response applications will need to be able to operate despite high humidity and potential for water splash and exposure. Temperatures may also be high at the same time. Additionally, logistics in those areas will be compromised where transportation capabilities and power infrastructure are likely limited.

For example, as described by the New Orleans District USACE, Operation Blue Roof is a priority mission managed by USACE for the Federal Emergency Management Agency. The purpose of Operation Blue Roof is to provide homeowners and permanently occupied rental properties in disaster areas with fiber-reinforced sheeting to cover the damaged roofs until arrangements can be made for permanent repairs. Operation Blue roof protects property, reduces temporary housing costs, and allows residents to remain in their homes while recovering from the storm.

This mission operates as a time sensitive priority mission in a logistically compromised environment. Timely response is required to best aid those affected by hurricane damages over large areas and should be improved by available means. Appropriate resources and crews need to be organized to distribute the roofing sheets and work to affected areas, however, that

organization can be delayed and inefficiently applied by lack of detailed information and logistics challenges introduced by storm damage. The operation requires survey and caller information to fully gauge the concentration of damages after storm fall, which can be skewed by response time of victims, who may be unable to contact USACE due to downed communications infrastructure. Response by roofing crews can be additionally stymied by blocked routes and flooded regions, adding delay to USACE response.

In such logistically compromised environments, external information systems can be used to quickly appraise the situation. Ariel reconnaissance can be rapidly deployed after storm dissipation to the damaged areas. This data can be generated by remote controlled drones with LiDAR and camera sensors. Affected regions and routes can be canvassed by drone scans and footage, which can be processed into 3D models and maps to provide accurate visual damage estimates for operation planners. Reconnaissance data can be aggregated to identify any obstacles to access of those regions, additionally feeding into situational awareness of the region, providing greater awareness for immediate and accurate disaster response.

Communications and power infrastructure in disaster areas is likely extensively damaged in regions of interest. Processing and aggregating the drone visual data into useful information needs to be done on external computing resources, but communications infrastructure cannot effectively support the raw data transfer required for timely appraisal of the situation. Edge computing resources with the capability to be deployed rapidly are required to support time sensitive response to disasters, with data processing in the environments lacking power and communications capabilities.

2.7.3 Mission Set Conclusion

Translating the data into useful information for those applications and scenarios requires significant computing resources to process data in timeframes acceptable. While research is dedicated to translating the data into useable field intelligence, the additional constraints placed by mission scenarios may prevent the data from being translated in a timely fashion via traditional high-performance computing (HPC) resources. For advanced sensor use, infrastructure conditions in the environment or external interference may prevent communication with cloud-based resources, necessitating the deployment of edge computing resources to those environments. The challenge lies in delivering HPC resources in the timeframe necessary for the mission, to already harsh environments with the added logistics obstacles. Edge computing resources must be able to deploy and operate in austere environments without requiring construction of dedicated facilities and high levels of logistics support.

2.8 Deployable Edge Computing Background

Examples of ruggedized deployable edge computing resources exist in certain configurations, which can be deployed to support computing functions in edge environments. Commercial solutions host a range of cloud/edge services in various forms of ruggedized edge appliances to deployable data centers in a scale of escalating logistics requirements for deployment. Microsoft hosts the Azure stack edge and cloud appliance products, a range of computing resources which can deliver Azure stack services to varying environments and applications, however, the combination of integrated cooling and environment protection is limited to the full-size data center.

Rugged Edge Portfolio

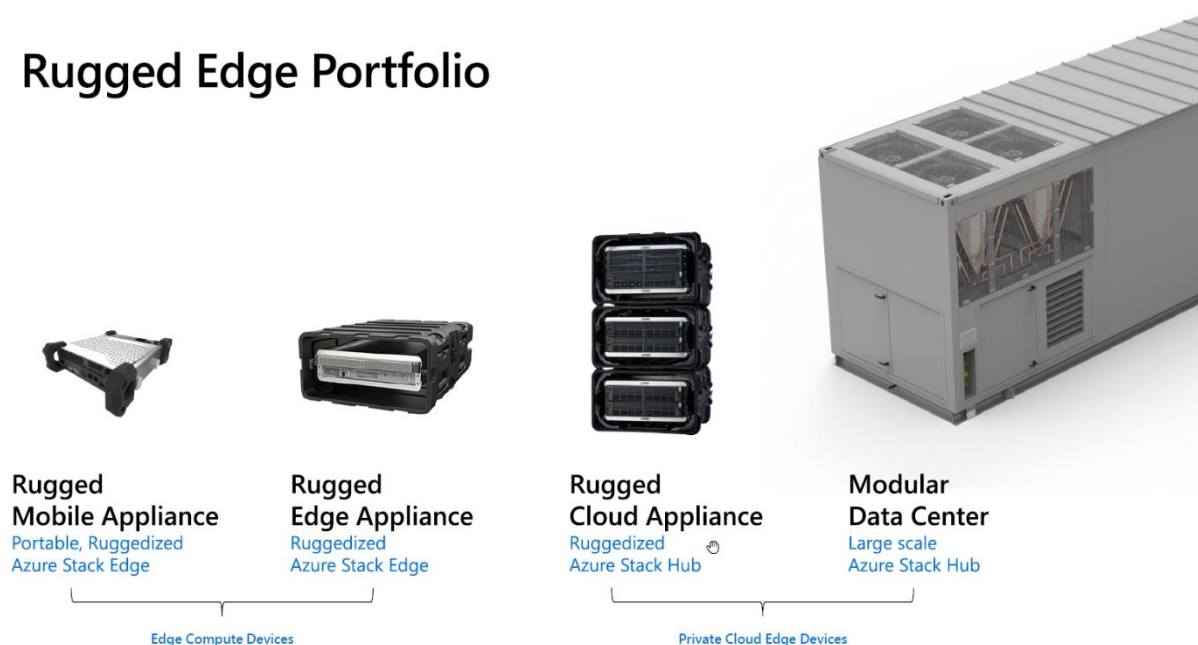


Figure 2.8 Existing ruggedized edge computing resources.

Microsoft Azure stack ruggedized product portfolio, providing Azure stack services in various configurations. Small deployable devices are ruggedized but not have cooling systems attached.

The product portfolio is representative of the current HPC/Cloud/Edge ecosystem. The smallest devices (Rugged Mobile Appliances) provide the Azure stack edge in a portable fashion with great transportability, although with limited computing power and data connections due to its lightweight design and size, 10.5" x 8.5" x 2.5" at 7 lbs. It hosts a 16 core Intel Xeon-D CPU, 24 vCPUs, 32 GB usable RAM and 750 GB SSD space, 2x 10Gbe SFP+, 2x 1Gbe RJ45 and Wi-Fi network interfaces, with battery and an AC power supply. The rugged edge appliance is also ruggedized for transportation, more adept in computing power as a 1U rack mount server, the total appliance comprising a single or four nodes with UPS system. This unit has 2 x 12 core Intel Xeon Silver 42124 CPUs, 128 GB of RAM, 8 TB flash storage, up to 2 NVIDIA Tesla T4 GPUs, 4x 25Gbe SFP+, 2x 1Gbe network interfaces. Package size is 31.5" x 23.8" x 11.9" at 93

lbs., or 179.23 lbs. Both edge appliances are configured for ruggedized transportation, however, rely on ambient air for cooling and do not have dust and water splash protection.





	Tactical Edge Appliance (TEA) Single Node with UPS	Tactical Mobile Appliance (TMA) <small>*Subject to Change/Pending Final Commercial Product Finalization</small>	Tactical Cloud Appliance (TCA)	Tactical Data Center (TDC)
				
CPU	Single Node – 20 CPU cores Four Node – 80 CPU cores	*16 CPU cores	High – 284 vCPU cores Low – 184 vCPU cores	Storage Optimized – 3,252 vCPU cores Compute Optimized – 44,022 vCPU cores
Memory	Single Node – 256 GB Four Node – 1024 GB	*48 GB	High – 1,037 GB Low – 547 GB	Storage Optimized – 24,291 GB Compute Optimized – 44,022 GB
Storage	Single Node – 53.8 TB Four Node – 215 TB	*16 TB	High – 34.2 TB Low – 15.4 TB	Storage Optimized – 10,168 TB Compute Optimized – 2,337 TB
Acceleration	Single Node – 1 FPGA Four Node – 4 FPGA	*2 VPU	N/A	Storage Optimized – 384 vGPU cores Compute Optimized – N/A
Size/Weight	Single Node with UPS – 31.5" L x 23.8" W x 11.9" H – Weight 92.93 lbs Four Node – 31.5" L x 23.8" W x 15.4" H – Weight 179.23 lbs <i>*Other SKUs available on request</i>	10.5" L x 8.5" W x 2.5" H Weight - < 7 lbs	Compute – 31.5" L x 23.8" W x 15.33" H – Weight 120 lbs (qty: 2) Networking – 31.5" L x 23.8" W x 17.12" H – Weight 145 lbs	40' L x 8' W x 10' H – Weight <48,000 lbs
Optional Configurations	Pre-Heater UPS	*73 WHr removable battery	Pre-Heater High or Low Configurations	Compute or Storage Optimized Configurations

Figure 2.9 USACE Azure Stack ruggedized edge devices.

Specifications of commercial ruggedized edge devices for the range of Azure Stack services. Updated specifications are described previously for the mobile appliance. Rugged, modular and deployable terms are interchangeable for this discussion.

The larger deployable appliances are the rugged cloud appliances and modular data centers, serving a central cloud functionality providing the Azure Stack Hub, while being transportable to edge locations. These devices are substantial in computing power and network functionality. The weight and size of these devices increases, constituting a full server rack of computing power, and finally a full containerized data center of computing power.

The spectrum of commercial products reveals limitations in the edge capabilities of ruggedized computing. Only the largest deployable unit, the data center, features integrated

cooling and containerization from the environment. As a 48,000 lbs. device, it lacks mobility during operation and ease of transportation. The smaller cloud, edge and mobile appliances are far more easily transported, however, lack protection from the environment and integrated cooling during operation. Commercial devices do not have the combined features of mobility and protection during operation to deliver high performance computing resources to edge environments with limited logistics capabilities.

2.9 RECON Established Requirements

In order to expand the use of advanced sensor technology to edge environments, computing resources deployed to those mission spaces need to be capable of operating in those environments. These capabilities form the overall design requirements and objectives of the RECON project. The advantages of edge computing can be leveraged in circumstances where physical and network logistics prohibit use of standard cloud HPC resources. Deployment of hardened, small scale, and mobile computing systems to edge environment conditions provides an opportunity to deliver computing capabilities for time sensitive and unconventional applications.

2.9.1 Objective

Therefore, the objective of the RECON project is to develop a system to house and operate a computing system capable of operating as an edge node for various computing applications. The system needs to be field deployable, and resilient to shock, high ambient temperature, dust and water ingress.

2.9.2 Requirements

RECON objectives, when considered under potential USACE applications in edge environments gain the following requirements:

1. The system must gain resistance to shock and vibration, enabling the transport of the NVIDIA DGX-1 GPU station and supporting components in a ruggedized case, for conventional ground and vehicle transportation.
2. The system must also deliver integrated closed-circuit cooling, preventing dust and water ingress, capable of cooling all computing components at full processing load, in external ambient temperatures of 115°F, 20°F above the preferred operating limit temperature of the NVIDIA DGX-1 (95°F).
3. Modularity of computing components, enabling ease of operation, protection, transportability and IP64 capable sealed transportation separate from the cooling system is required.
4. System maneuverability through 3-foot-wide doorways while deployed is required.
5. The system must operate using either provided on-site shore power and generator power.
6. Overall, the system is required to deliver high performance computing capabilities to the edge, with attributes of survivability, mobility and maneuverability.

2.9.3 Research

Considering the applications, objective and requirements above, is it possible to construct a ruggedized edge computing node with the necessary attributes of survivability, mobility and maneuverability, using commercially available computing, cooling and ruggedized packaging, to deploy HPC capabilities to the edge, beyond conventional HPC logistics?

CHAPTER III

CONCEPT DESIGN DEVELOPMENT

3.1 Introduction

Initial design concepts were explored to understand option for three basic points of design: protection of operating computing components, cooling of computing components and the ruggedized packaging of entire assembly. In short, the ruggedization, cooling and framing design were considered. The size and level of computing is selected by the gaps in existing deployable edge computing technology, in a scale between processing power and mobility. Based on that determination, where the scale of computing components is known, the manner in which they can be cooled, packaged, and additionally operated and powered is explored through concept design. Ruggedized transportation of rack components is relatively well-established commercially, however when combined with the other necessary functionalities, the concept designs require additional solutions. Attributes of mobility, maneuverability and survivability are balanced with varied design focus and features of each approach.

In this case, components are required to be protected during transportation, while also providing cooling air circulation routes for adequate heat dissipation. Contingent to the packaging, the configuration of cooling needs to also fulfill transportability requirements while matching cooling capability to the heating capacity of full load computing. Configurations of these two factors, with the computing components, supporting hardware, and accessories, along

with user interaction requirements established throughout the development of this project, are explored in the concept designs and iterations.

In the development of the concept responses to the objectives and requirements, the design will form according to the response priority to the functional objectives above, achieving base requirements while also accommodating the functionality traits such as transportability and maneuverability. First, the basic solutions to the cooling and ruggedization must be established, before the mechanical design can be explored to achieve the rest of the objectives. These objectives will also change in scale of priority, also affecting design choices through the concepts and later prototype iterations. Design response and narration can be described according to the fulfillment of those trait objectives.

RECON Attributes and Objectives

MOBILITY	SURVIVABILITY	MANUEVERABILITY
<ul style="list-style-type: none">• Transportability• Minimal system dimensions• Minimal weight• Modular components	<ul style="list-style-type: none">• Ruggedization, shock and impact resistance• Resilient structure and framing• Environmental protection and insulation• Cooling capacity against high ambient temperature• Ingress protection against dust and water	<ul style="list-style-type: none">• Ease of operation and deployment• Access to controls• Transportability during operation• Flexible power management

Figure 3.1 RECON base attributes and objectives.

Functionality attributes are visually associated with design objectives. RECON requirements determine minimum functionality in each. Favorability of certain objectives over others results in different design approaches of RECON iterations.

3.2 Approach Overview

The design narration of RECON will follow the design responses of the formative components as they are considered against the use case, or operation in edge environments. The sizing and computing power of components is determined in response to the capabilities, or lack of, existing deployable HPC resources applicable to the mission space. Following the selection of computing components, the individual systems of ruggedization, cooling and framing configuration are explored, oriented towards attributes of survivability, mobility and maneuverability. Finally, the cohesive concept design iterations are reviewed for their approach to the requirements.

Concept Progression

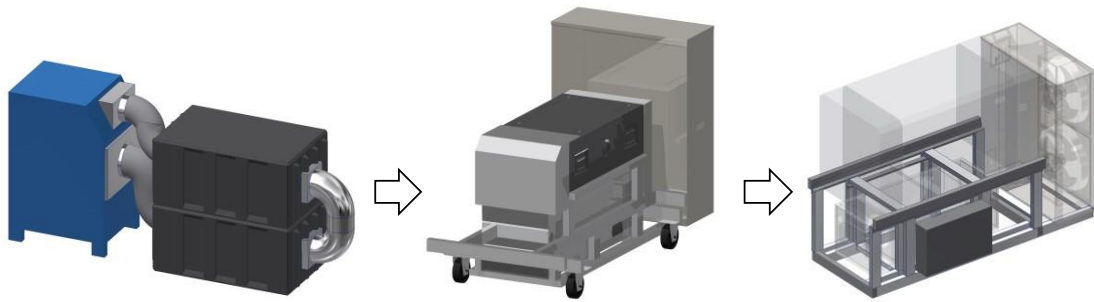


Figure 3.2 Concept design progression.

Progression of concept designs as individual component systems are selected towards fulfillment of RECON requirements.

3.3 Computing Component Selection

3.3.1 Component Objectives

Computing resources with the best performance in responding to various tasks and sensor processing needs require flexibility of processor types. GPU and CPU cores with as much processing power possible under packaging constraints are required. Multiple Unit rack systems, such as the previously mentioned Rugged Cloud Appliance, can house the GPU, CPU, memory, networking and power supply hardware required for HPC tasks in diverse environments. In edge scenarios where the logistics capabilities are limited, computing components need to be transportable with minimal equipment or machinery requirements, and should be transportable into existing structures, in anticipation to the example USACE mission set of operating in improvised disaster relief facilities. Multiple case or single case computing configurations weighing from 250 - 500 lbs. maintain transportability without lifting equipment. Ruggedized rack units, such as the 31.5" L x 23.8" W x 15.3" H Rugged Cloud Appliance maintain a packaging size able to fit through doorways with their maximum length and width dimensions,

indicating that stacked server rack components in ruggedized packaging can be transported into existing facilities. Computing compositions that combine powerful server rack CPU and GPU units, with the necessary supporting network, memory and power supply hardware into a ruggedized server rack, can provide HPC resources transportable to edge environments.

3.3.2 Computing Components

The primary components of each iteration of RECON constitute the high-performance computing stack, focused around supporting the NVIDIA DGX-1 and additionally the Dell 7920 Precision rack workstation, the main GPU and CPU processors. These units provide the computing versatility required for edge environments and are powerful computing units for their footprint. The computing components are the core components of which the RECON platform is built around, the core of its functionality and requirements. At peak computing load, the system will need to dissipate 4950 Watts of electrical load to support up to 1000 Teraflops of GPU processing power from the DGX-1 setup. The configuration supporting both the DGX-1 and Dell 7920 Precision Rack Workstation is 6050 Watts. In order to support a standalone HPC resource in an edge network environment, the system contains a hard drive unit, switch, power distribution and controller, selected as part of the initial design requirements. The total device footprint, cooling requirements and weight to compensate for can be accounted for in this initial configuration.

3.3.2.1 NVIDIA DGX-1

The NVIDIA DGX-1 GPU unit is equipped with 8 NVIDIA Tesla V100 cores, 32 GB per GPU, 40960 total NVIDIA CUDA cores and 5120 Tensor cores, handling large GPU based processing. System computing is handled with two 20-core Intel Xeon CPUs, specified as E5-

2698 v4 2.2 GHz. On board memory is a 512 GB DDR4 RDIMM, with cache of four 1.92 TB solid state drives. Four InfiniBand 100 Gpbs EDR and two 10 GbE connections handle the network connectivity and interaction. The total assembly is powered by four 1600-watt power supply units, overall providing for the 3500 W peak power consumption rating. The NVIDIA DGX is a rack mounted unit; however, it is unique in form factor that in its depth, at 34 inches, it is longer than most units. The DGX is 3U in height in rack mounting, dimensions are 34-inch depth x 17.5-inch width x 5.15-inch height. DGX-1 unit weight is 134 pounds. Cooling flow and component diagram are shown in NVIDIA graphic below, Figure 3.4. Temperatures of cores and overall unit intake are self-regulated by fan speed in the DGX unit, similar to other rack mount units.



Figure 3.3 NVIDIA DGX-1 front face.
3U height component.

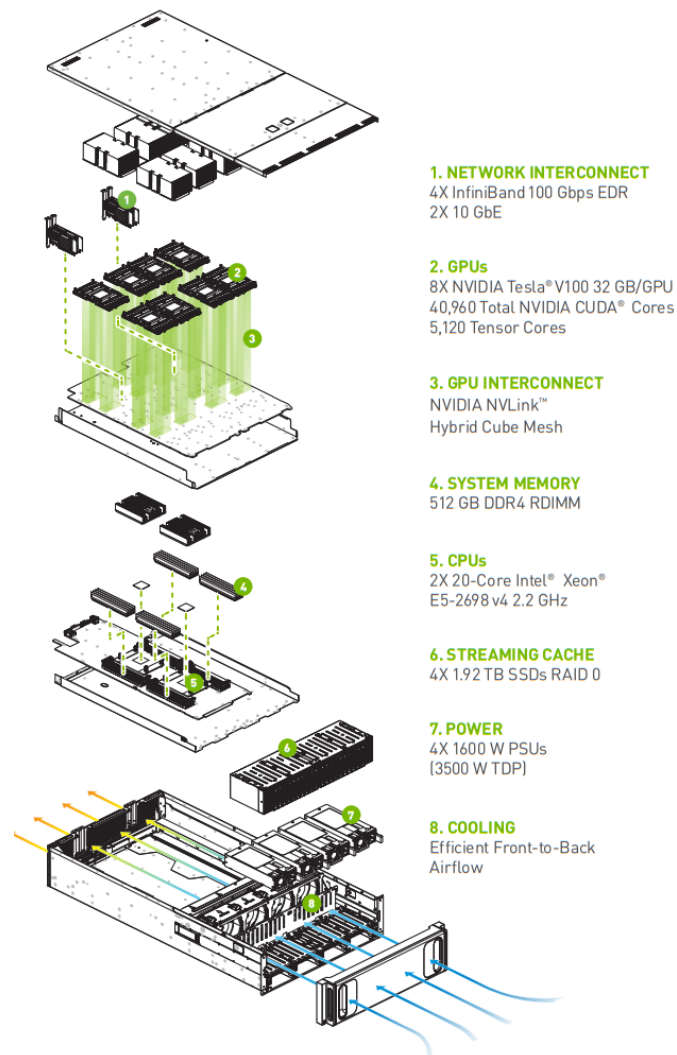


Figure 3.4 Diagram breakdown of Nvidia DGX-1 components.

3.3.2.2 Dell Precision 7920 Rack Workstation

An addition to the original processing components, the Dell Precision 7920 Rack workstation is the CPU focused processor for the system. It is a dual 28 core Xeon platform, supporting computing versatility of the system for applications requiring CPU based architectures. As an overall workstation, the Precision features the hardware interfaces seen on

tower computers, including hard drive, network, device ports and additional GPU ports. This unit is equipped with the Intel Xeon Gold 6240 CPU at 2.4 GHz, 128 gigabytes of RAM, and a Nvidia RTX 6000 graphics card. The Dell Precision rack unit is 2U in height, at 63 pounds. Dimensions are 28.17 inches depth by 19 inch width and 3.42 inches in height. Through cooling is handled by hot swappable fan units with two pairs of fans, dissipating a peak of 1100 watts at peak processing power.



Figure 3.5 Dell Precision 7920 rack workstation front face.
3U height.



Figure 3.6 Internal component view of Dell Precision 7920 rack workstation.

3.3.2.3 DDN SFA220NV

Dedicated data storage is handled by the DDN SFA220NV hard drive unit. The Data Direct Networks storage unit “Storage Fusion Architecture” is selected for its high data transfer rate of 33 GB per second, in addition to the storage unit’s file storage structure working to support the potential deep learning, and artificial intelligence applications of the DGX-1. The flash storage system is expected to use 150 Watts at writing capacity. Rack unit is 2U in height, 3.42 inches, weighing 90 pounds.



Figure 3.7 DDN SFA200NV hard drive unit front view.

3.3.2.4 Mellanox InfiniBand SB7700

Data transfer is managed by the Mellanox InfiniBand SB7700. It hosts 36 EDR 100 gigabit per second ports, with 7.2 terabits per second of aggregate switch throughput. These speeds are achieved with 90 nano seconds switch latency. Internal heat is managed by a redundant and hot swappable fan system, in addition to redundant power supplies. The unit dimensions are 27.25 inches depth, by 15.875 inches width and 1.75 inches height as a 1U unit. The switch weighs 30 pounds and will use 1300 Watts at peak performance.



Figure 3.8 Mellanox InfiniBand SB7700 36 port switch.

Front face shown without rack mount hardware, 1U height.

3.3.2.5 Raritan IX7 Power Controller

Internal power for the system is routed through a Raritan IX7 smart power controller. The system monitors power draw and amperage through each channel, providing a data stream out to alert of high wattage conditions or amperage conditions. The system supports temperature and humidity sensing with I2C serial wire sensors. Using application interfaces, the data can be passed through and logged to record power usage patterns and air temperature responses of components in proximity to the PDU in the rack mount. The unit is not a full depth rack unit, taking space in one side of the mounting solution.

3.3.2.6 Computing Pipeline

The overall computing pipeline is optimized for the Nvidia DGX, as the Mellanox InfiniBand switch, and the DDN hard drive are optimized to work together to support DGX GPU functionality and rapid data management. Additional components can be added to support functionality requirements, such as network and power regulation. All components are rack mounted units, with one way directed fan cooling as needed to operate in a typical server rack environment.

3.4 Component Ruggedization

3.4.1 Objectives

Ruggedization of the computing stack needs to address the design requirements and primarily achieve the attribute of survivability. In addition, design of component ruggedization needs to enable attributes of mobility and maneuverability for itself and the system in order to achieve cohesive function. Commercially available server rack ruggedization and packaging provide multiple options for system design, however, must accommodate the unusual mission

objective. Component ruggedization and packaging option are reviewed up through the concept configurations of RECON until the final design selection.

3.4.2 Requirements

Requirements for ruggedization of the computing components ensure survivability for anticipated environments. The computing stack must be transportable without lifting machinery, insulated from the operating environment, resistant to dust and water splash ingress, and be able to circulate cool air flow through the components. Transportation of the computing equipment to edge scenarios introduces shock and vibration unusual to the equipment. Ruggedization of the components requires that they are securely mounted on a dampened rack system, tuned for the weight of the total system. Computer rack encasement must also be capable of withstanding direct impact expected during transportation, case materials need to be able to withstand damage. Ultimately, the modularity of the computing system needs to enable for separate transportation from cooling assembly and should be portable without lifting equipment.

3.4.3 Approach

Review of commercial rack mount ruggedized transportation reveals multiple possible configurations that address requirements. Dampening methods, case materials, internal geometry, and external features are considered. The dampening methods must be tuned to the weight of the internal computing components which have a total weight of 327 lbs. The largest component is the NVIDIA DGX-1, with a depth of 34”, and the total height of the computing stack is 8U. The depth of the DGX-1 exceeds the typical depth of rack components and clearance for intake and exhaust of the rack components must also be considered.

3.4.4 Medium Duty Polymer Rack Mount Cases

Initial concepts for ruggedization considered the Pelican V-Series and Classic Rack Mount Cases. Classic-V cases feature 33” rack depth, 19” rack width for #10-32 or M6 square hole clip nuts, with 2.5” and 5.25” front and back lid depth, supporting up to 170 lbs. payload with 9U rack height. Vibration and shock dampening occurs over 1.4” of internal sway space. They feature edge casters, stainless steel handles and molded stacking ribs. Cylindrical shocks are connected to the case walls, isolating shock and vibration for medium duty applications. Case walls are made of rotomolded polypropylene, offering protection equivalent to the duty. Heavy duty cases or other case series by Pelican are not applicable, as they did not offer adequate rack depth for the DGX-1 in combination with adequate payload capacity and rack height.



Figure 3.9 Pelican rack mount case options.

Pelican case concept options explored for ruggedization, stackable and medium duty rack case protection.

Applying Pelican or similar commercial cases to the RECON computing stack could be accomplished by splitting computing components across two cases, due to payload capacity and appropriate rack height combinations. Front and back lid clearance from the rack is 2.5” and 5.25”, offering limited space for air circulation. In order to support operation of the components during operation, data communications will need to be passed between cases through bulkhead connections, and air circulation needs to be ducted to both cases, increasing operational complexity, reducing maneuverability. Using multiple cases will increase the modularity of the system, however, enables easier transportation of the components, lending to system transportability and mobility. Polypropylene case wall material achieves component protection, however, is susceptible to direct damage and abrasion, which additionally weakens in high temperature environments, which may warp under stacked loads with the RECON computing stack, limiting the survivability of this solution.

3.4.5 Heavy Duty Aluminum Rack Mount Cases

In pursuit of ruggedized cases with higher maneuverability and survivability, larger customized commercial cases were considered. This approach selected cases which can house the entire computing stack, offering impact and vibration dampening with front and back clearance for air flow. Impact Cases offers portable Mil-Spec aluminum cases for server rack ruggedization, focused on the ruggedization necessary for operation in edge environments. These cases feature aluminum welded and riveted body construction, cylindrical dampers and steel square hole steel server rack frame which support up to 500 lbs. payload. Internal features are designed to withstand high stress conditions, resisting deformation under load and maintaining watertight seals. External ribs provide case rigidity and impact protection for metal hardware,

which function as hand holds, strapping points and latches on the lids and body. Extended lid depth is available for integrated cooling inlet and outlets, with features for ducted cooling.



Figure 3.10 Target Impact Case with standard ruggedization features.

3.4.6 Ruggedization Selection

A customized Target Impact Case was selected for the RECON ruggedization and packaging. Focusing on survivability, this case selection also best achieved aspects of mobility and maneuverability for anticipated methods of cooling and on-site operation. This heavy-duty case focuses on accommodating the RECON computing stack with full mounting solutions and internal clearance. The body is made of 0.125" 5052-H3 aluminum welded panels, with rivetted body ribs for impact deformation resistance. The body length is 40", extended to accommodate the Nvidia DGX-1, housing the 12U rack height and 34" depth steel server rack frame. Server rack shock absorption payload is rated for 251-350 pounds for the 327 lbs. RECON stack, using

8 cylindrical dampeners connecting the frame to body. Overall case dimensions are 58” length, 28” width, by 35” height. The overall case protection features will reduce vibrations and dampen large impacts to the computing stack, enabling operation and transportation to the intended environments.

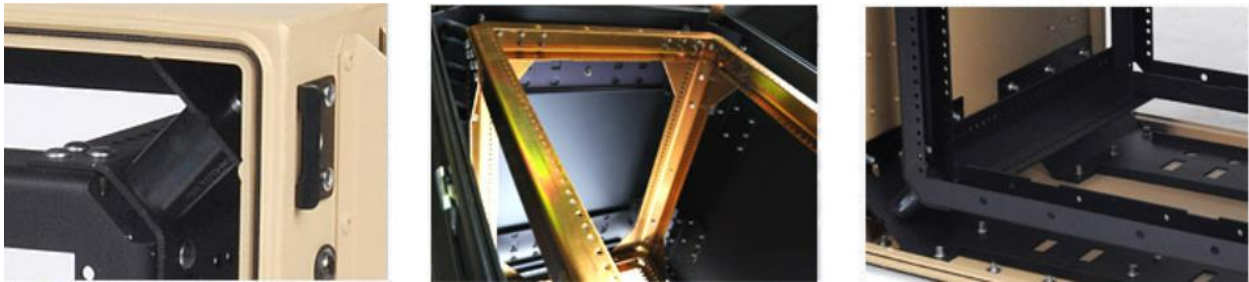


Figure 3.11 Impact Target case internal features.

Highlighted internal features, left, include the double gasket seal and seam welded construction securing the dampening structures to the body. Center, the frame rack system is made of U-channel welded and single pieces with reinforced framing in anticipation of high load, secured against vibrations. On the right, internal load spreaders for the server rack prevent distortion and ensure case rigidity up to 500 lbs. payload



Figure 3.12 Target Impact Case.

Target Impact case with custom features designed for heavy duty protection of the RECON computing stack. Ruggedization of components and functions aided by hinged panel cover of bulkhead connections in the center of the front face, rigid structure and handling hardware.

Cooling and environmental protection is accounted for using 8" front and back lids with 6" x 22" cutouts on the bottom face to provide A/C ducting integration and space for cool air distribution, ultimately providing unrestricted airflow for the DGX-1. The case achieves IP64 ingress protection rating when the cutouts are sealed. A customized recessed panel hosts all bulkhead data connections to maintain that rating, with a hinged panel cover for transportation. Additionally, the inside walls are lined with 1" thick insulation, reducing heat transfer from the external metal shell.



Figure 3.13 Impact Target case customized lid features.

Cutout features on case lids, left, are sealable for transport when internal airflow is not required.

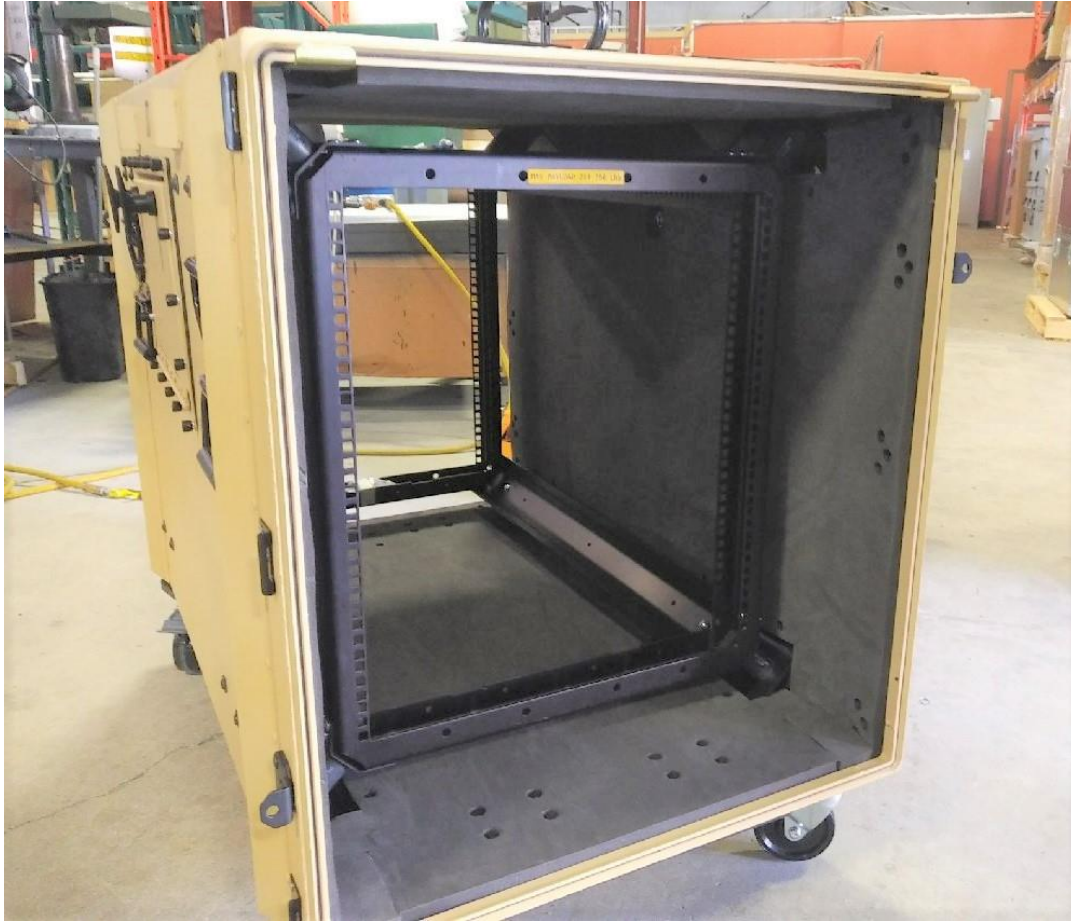


Figure 3.14 Impact Target case internal rack features after insulation.

Considerations for the maneuverability and transportability are included in the removal and addition of features. Caster wheels can be locked into place beneath the case or removed and stowed within the case lid, supporting modular transportation separate from cooling units and attached equipment. The forklift features were removed to maintain flat geometry below the case in anticipation of cool air ducting design, in addition to reducing the height of the overall unit.

3.4.7 Ruggedization Conclusion

The Target Impact case was selected as the ruggedization option for the RECON computing stack. Its heavy-duty approach to the containerization of the components, offering adequate dampening and sway clearance for the payload, versatility with removable lids and caster wheels, insulation from the ambient environment, and protection during transportation and operation provides good reason to integrate the unit into system design. Successful evaluation of this ruggedization approach will ultimately be determined by the survival of the computing components in the target environments, where the system ruggedization also allows for ease of transportation, cooling and operation.



Figure 3.15 RECON ruggedized computing stack.

RECON computing stack front face, installed within Impact Target case. In addition to the main processing stack, the 12U rack height allows for the addition of a UPS unit at the bottom, and an additional network switch, top.

3.5 Cooling System Selection and Concept Integration

3.5.1 Cooling System Requirements

Ruggedized HPC systems require cooling systems to operate in edge environments. Ambient air can be damaging to RECON components via dust, water and temperature, requiring closed circuit air circulation through the components. The cooling system is required to provide enough cooling power for the computing stack while under full processing load, while enabling the mobility of the entire system, maintaining roughly the same logistical footprint. Commercial systems were reviewed and selected for integration.

Estimated total heat output of the NVIDIA DGX-1 primary computing stack (including the Mellanox InfiniBand Switch, DDN SFA220NV Hard drive, and power distribution unit) totals 4950 Watts at full load, or 16804 BTU/HR, equaling the cooling capacity required. Simultaneous function of both the NVIDIA DGX-1 and the Dell 7920 Precision bring the estimated maximum power dissipation to 6050 Watts, or 20557 BTU/HR. Computing operations tend to use only one of the two units for their primary processing operations. In addition to heat conducted into the system from ambient temperatures, the system must also be able to dissipate the maximum heat load into 115°F environments, maintaining the operating air temperature between 50°F and 95°F, safe for the server rack components to regulate themselves within. Cooling requirements were also adjusted for an estimated heat gain through the case and stand in air handler, increasing the DGX-1 primary and DGX-1 with Dell 7920 Precision heat loads to 19376 BTU/HR and 23129 BTU/HR respectively at 115°F ambient temperature at an operating temperature of 95°F.

Estimated heat gain into system is estimated by assuming a heat resistance value across the surface area of the selected Impact ruggedized case and an equal surface area for a stand in

air handling unit. With a height of 2.92 ft, length of 4.83 ft, width of 2.33 ft, the surface area can be applied across the temperature differential at peak temperature operation over the assumed thermal resistivity.

$$\begin{aligned} \text{SurfaceArea} &= 2(\text{Height} \times \text{Width} + \text{Height} \times \text{Length} + \text{Width} \times \text{Length}) \\ &= 64.3 \text{ ft}^2 \end{aligned} \quad (3.1)$$

$$\Delta T = (115^\circ\text{F}) - (95^\circ\text{F}) = 20 \Delta^\circ\text{F} \quad (3.2)$$

$$R = 1 \frac{\Delta^\circ\text{F ft}^2}{\frac{\text{BTU}}{\text{hr}}} \quad (3.3)$$

$$Q = \frac{\text{SurfaceArea}}{R} \Delta T = 1286 \frac{\text{BTU}}{\text{hr}} \quad (3.4)$$

$$2 \left(1286 \frac{\text{BTU}}{\text{hr}} \right) = 2572 \frac{\text{BTU}}{\text{hr}} \quad (3.5)$$

The surface area is calculated across the ruggedized case dimensions. The temperature differential between the maximum ambient temperature and the maximum preferred closed circuit air temperature for the DGX-1, or the return air temperature, is found to be 20°F. A low thermal resistivity value is assumed for the case surface area, compared to typical thermal resistivity values afforded by insulation 1” thick or more. Heat transfer is estimated for those values, and the overall value is doubled to account for estimated air handling unit and framing heat gain by conduction. The heat gain through the case and ducting surfaces is increased when the closed-circuit temperatures are below the operating limit of 95°F. At lower the lower bounds of expected temperature operation of 65°F, conductive heat gain would be 6430 BTU/HR. Following, heat gain at 75°F is 5144 BTU/HR, and 85°F is 3858 BTU/HR. Total heat gain to be dissipated by the cooling system in 115°F ambient by the condenser and computing case and

ducting would be 26987 BTU/HR for a target internal temperature of 65°F, and similarly, 25701 BTU/HR for 75°F, and 24415 BTU/HR for 85°F. Cooling system selection and overall system design should account for the range of heat gain, in addition to reducing that heat gain with more substantial insulation than the assumed R-1 rating applied in the estimates, and an overall reduction in exposed surface area.

3.5.2 Approach

Commercial solutions to the required cooling application must simultaneously have enough cooling capacity while operating under a limited logistics footprint, fulfilling the mobility requirements of at least being able to move through three-foot-wide doorways. Portable rooms coolers, split unit A/C systems, integrated cooling and server rack coolers were considered across different computing system ruggedization concepts.

3.5.3 Integrated Coolers

As an immediate choice by proximity, existing integrated Impact coolers for the ruggedized case were considered. These units provided closed circuit climate control with low vibration, filtered intakes and a rain baffle for compact integrated cooling directly mounted to the case, providing target levels of compactness for transportation, modularity, ease of operation and survivability. These integrated solutions and other similar market solutions, however, did not accommodate for the HPC cooling requirements of this effort, limited to 6000 BTU/HR at 125°F, roughly a third of the required cooling capacity required.



Figure 3.16 Commercial Impact integrated case cooler.

Integrated Impact case removable chiller, above the server rack case, provides closed circuit air into cutouts on the case lids. Cooling capacity is limited to 6000 BTU/HR at 125°F, inadequate for RECON requirements.

3.5.4 Portable Coolers

Larger capacity chillers are required to operate the RECON computing stack. Separate commercial chillers need to be integrated to provide adequate cooling capacity to the RECON stack, but also need to maintain a footprint capable of maneuvering through doors and hallways of existing facilities. Portable cooling units provide this maneuverability. High-capacity portable chiller, such as the Tripp Lite portable cooling unit, for example provides 24000 BTU/HR cooling capacity at 240V, dimensioned 20.5" width, 22" depth, and 38" height, not including

ducting, at 175 lbs. the chiller is compact and can use directed ducting to exhaust heat outside the environment and directly into the case.



Figure 3.17 Portable air conditioning unit.

Example commercial portable cooling unit air conditioner, by Tripp Lite, provides 24000 BTU/HR cooling capacity. Compact body requires ducting solutions to extend cooling to computing case and direct waste heat external to operating environment. The right shows ducting to server racks.

3.5.5 Portable Cooler Concept

This chiller would need to be ruggedized itself to fit mobility and survivability requirements to match the computing case. In high ambient temperature conditions, the unit may fall short in cooling capacity to the full load RECON computing stack. In a modular, standalone concept with the portable cooler providing ducted cooling alongside the computing cases, flexible ducting could connect to ducting plenums, adapting air flow to the cases. This solution would require assembly on site, reducing maneuverability of a deployed system, and exposed

ducting and duct plenums outside the case and cooler bodies can easily suffer damage, reducing survivability and reliability on site. Considering a concept assembly using Pelican cases and this portable cooler, the assembly would be highly modular and transportable, at the cost of deployment complexity, if the case, ducting and cooler lack the hardware to integrate them together.

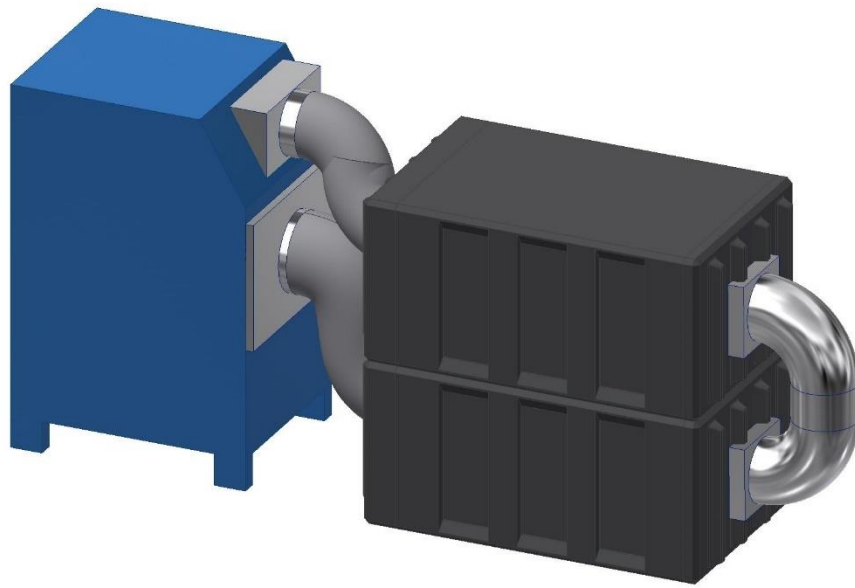


Figure 3.18 Portable cooler with polymer case concept.

Concept assembly of portable cooling unit providing closed circuit cooling to two Pelican cases, in series circulating air through installed ducting. This concept emphasizes modular design and transportability, over survivability and on-site maneuverability.

3.5.6 In-Row Server Coolers

Larger capacity coolers are needed to fulfill the cooling requirements of the full RECON computing stack in adverse environments. Considering the existing footprint of the Impact case,

commercial server rack coolers were considered for their increased cooling capacity. Server rack coolers operate alongside server racks, sharing the same footprint, and can be potentially integrated for mobile function. Commercial examples of server rack could then be integrated into the existing case ruggedization, providing mechanically mated closed circuit cooling directly to the computing components. Ruggedized ducting structure can be used to integrate the systems and introduce hardware that increases system mobility. In row commercial air conditioning units considered can cool higher capacities, the example Tripp Lite unit can cool 33000 BTU/HR, adequate to cool 20000 BTU/HR even under derated conditions. In row units house both evaporator and condenser functions within their body, dumping waste heat into the immediate environment. Body dimensions are 23.6" width, 43" depth and 78" in height at 560 lbs.



Figure 3.19 In-row server cooler commercial unit.

Tripp Lite example commercial in-row air conditioning unit considered for integration to RECON cooling, capable of 33000 BTU/HR cooling capacity. Side view, on right, shows both evaporator and condenser with fan units.

In row air conditioners with combined condenser and evaporator systems in the main body result in too large a system for transportability requirements. If the combined air conditioning unit was mounted above the Impact case with the RECON computing stack, similar to the existing Impact integrated cooler, the combined height would be 6.5'. Not including any additional mounting or ducting structures and height added from wheels, the basic dimensions would limit the overall assembly's transportability into facilities and transportation vehicles. Additionally, the total weight of both the computing stack and the air conditioner would near 1000 lbs. Over just four caster wheels, system stability over anything other than flat ground would be limited.

3.5.7 Split Server Coolers

In order to create a mobile integrated ruggedized HPC computing stack and air conditioner, smaller air handling units are required. Combined evaporator and condenser units are too large to integrate with the computing case. Split air conditioning units were selected in the concept that the condenser unit could be mated to the computing case, while the evaporator unit could be separately mounted. This approach maintains functional transportability by distributing the air conditioning system across an acceptable footprint. Framing systems can be introduced to add mobility and maneuverability to the more integrated system. The high cooling capacity of server rack coolers can be applied with a limited system modularity if their size allows for a transportable integrated evaporator and computing stack, with a trailing condenser unit.

3.5.8 Split Server Cooler Concept

The commercial solution considered for a split unit server cooler is the Ice Qube SC Series Mini Split Air Conditioner. The system was selected for its form factor in combination to its cooling capacity, made to deliver configurable cool airflow in server rack applications. The cooler is designed to deliver closed circuit cooling within the footprint of a 19" standard server rack, where the controls, evaporator, and air handling components are housed within a 16" height, 17" width by 45" depth body, and the condenser, 46" height, 36" width, by 14" depth. The evaporator unit weighs 96.8 lbs., and the condenser weighs 185 lbs., a total of 281.8 lbs. This unit weighs 278 lbs. less than the previously examined in row server cooler. The selected unit is capable of 27000 BTU/HR cooling capacity, depending on the return air temperature within the closed air circuit and the ambient temperature about the condenser heat exchanger.



Figure 3.20 Split cooler design concept with medium duty case and integrated frame.

Cooling design concept integrating an example Pelican case computing stack above a compact split server rack air conditioner, modeled after dimensions of the 27000 BTU/HR Ice Qube unit. Ducting and framing hardware mate the two units and ensure closed circuit cooling, while mounting the condenser unit behind the main body.

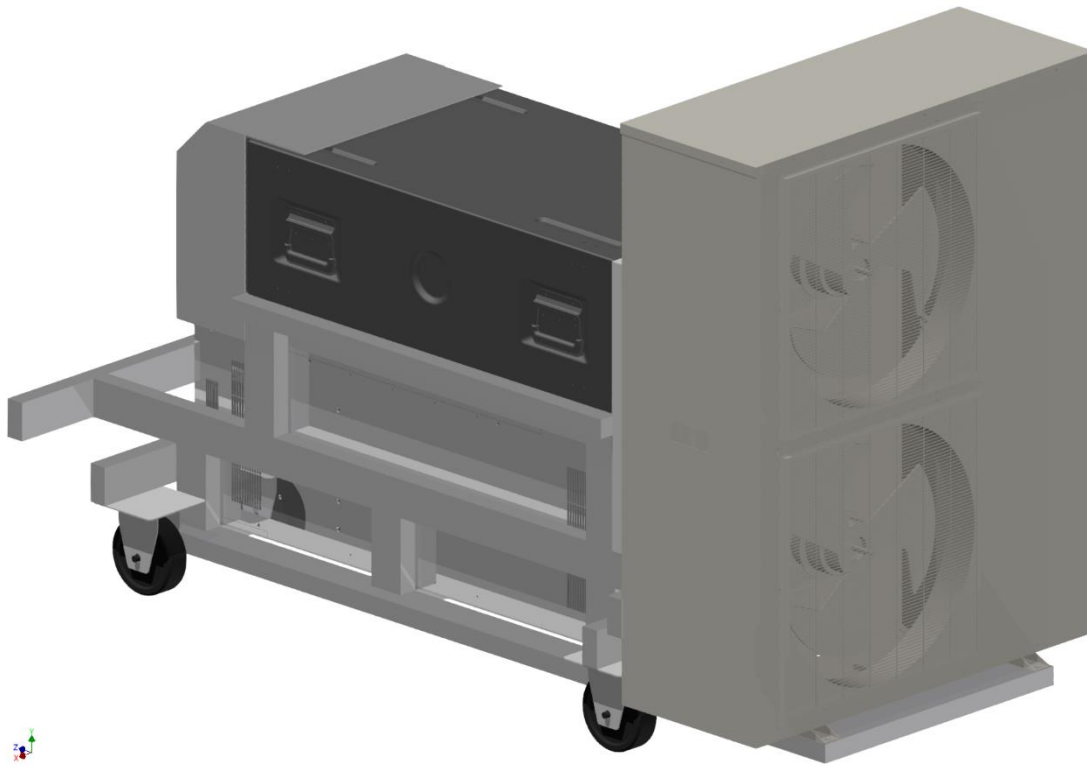


Figure 3.21 Split cooler design concept rear isometric view.

Additional view of the integrated computing stack and evaporator layout. Feasibility of condenser orientation in regard to mobility is evaluated in further iterations of concept design.

3.5.9 Cooling System Selection

The Ice Qube mini split air conditioner with 27000 BTU/HR capacity was selected for full concept design. This server cooler can adequately cool full load processing of RECON computing components at the requirement ambient temperatures of 115°F for the energy dissipated purely by the full processing load of computing components, 20557 BTU/HR against the cooling capacity, 21500 BTU/HR. That cooling capacity is exceeded when estimated heat gain into the system is considered, raising the heat load requirements to 23129 BTU/HR.

Increasing conductive heat gain, where heat gain increases for cooler internal circuit temperatures, at 24415 BTU/HR for 85°F, 25701 BTU/HR at 75°F, and 26987 BTU/HR for 65°F. Estimated heat gain would exceed capabilities for the Ice Qube system if placed in 115°F ambient external temperatures. Considering the physical assumption and parameters placed on the estimate for heat gain through the case and cooling system surfaces, with low insulation thermal resistivity of R-1 and a surface area calculation doubling the exposed surface area of the computing case, design and operation parameters can be adjusted to account for the estimated heat gain. Achieving higher R values across surfaces, reducing exposed surface area by combined air handler to computing case configuration, and ducting around the smaller footprint of the IQ27000V compared to the computing case will easily reduce heat gain into the system into feasible ranges for evaluation. Additionally, it was determined that the full load of processing across all components would not be sustained in practice, however, and the potential for short periods of heat gain within the system was accepted against the requirements. Thermal cycling, occurring when the split system cycles on and off in cooling according to temperature thresholds, was the selected mode of operation natural to the cooling system. In application, this would cool the closed-circuit computing components down to the lower threshold temperature until the off cycle, where the ambient heat gain through surfaces and heat generated by continued component operation would heat the circuit until the control threshold is met or maximum operating temperature of 95°F is reached. If full cooling capacity against external ambient temperatures was reached for a constant processing load, the cooling system would no longer cycle, and the internal temperature would rise until conductive heat gain reduced to balance the system with respect to energy exchange and temperature. On the upper bound of internal circuit temperature at 95°F, control processes for the NVIDIA DGX-1 GPU cores or Dell Precision

7920 CPU cores would automatically down regulate their processing speed according to their core temperature regulation controls. Therefore, the cooling system should be able to dissipate the heat output from the computing components, otherwise reaching a balanced state at higher temperatures, until computing components automatically down regulate heat output under exceptional use cases. The Ice Qube 27000 BTU/HR capacity Server Rack Cooler IQ27000V (IQ8000SC condenser and IQ8000SE evaporator) was selected as the prototype cooling system for the RECON computing stack.

The physical footprint of the cooling system works under the integrated cooling system and ducting concept design. Physical requirements place the design restrictions to essentially be able to fit through facility doorways. Air handler dimensions of the split unit are smaller than the footprint of the computing case, fitting vertically overhead within the contour of the computing case. Server rack cases and computing components capable of achieving HPC capabilities maintain the same general footprint, otherwise varying in rack height for the addition or removal of components. Potential computing upgrades such as the DGX-1 to the DGX-2 require additional rack height and cooling power, and can be achieved under this component selection, where the computing case is fabricated with additional rack capacity, and the IQ27000V is replaced with the IQ34000V for its increased cooling capacity. Component selection here maintains the overall intended physical size, requiring roughly the same logistics consideration as existing Cloud Tactical Appliances, while allowing the standalone, field-deployable capabilities required. Optimization of the integrated design with the IQ27000V for transportability or maneuverability will depend on the design response for the chosen priorities.



Figure 3.22 Ice Qube split server cooler condenser and evaporator.

Refrigerant line hookups and control panel are seen on the right of the evaporator, bottom. The condenser unit for the 27000 BTU/HR unit is a double stacked unit compared to the unit shown.

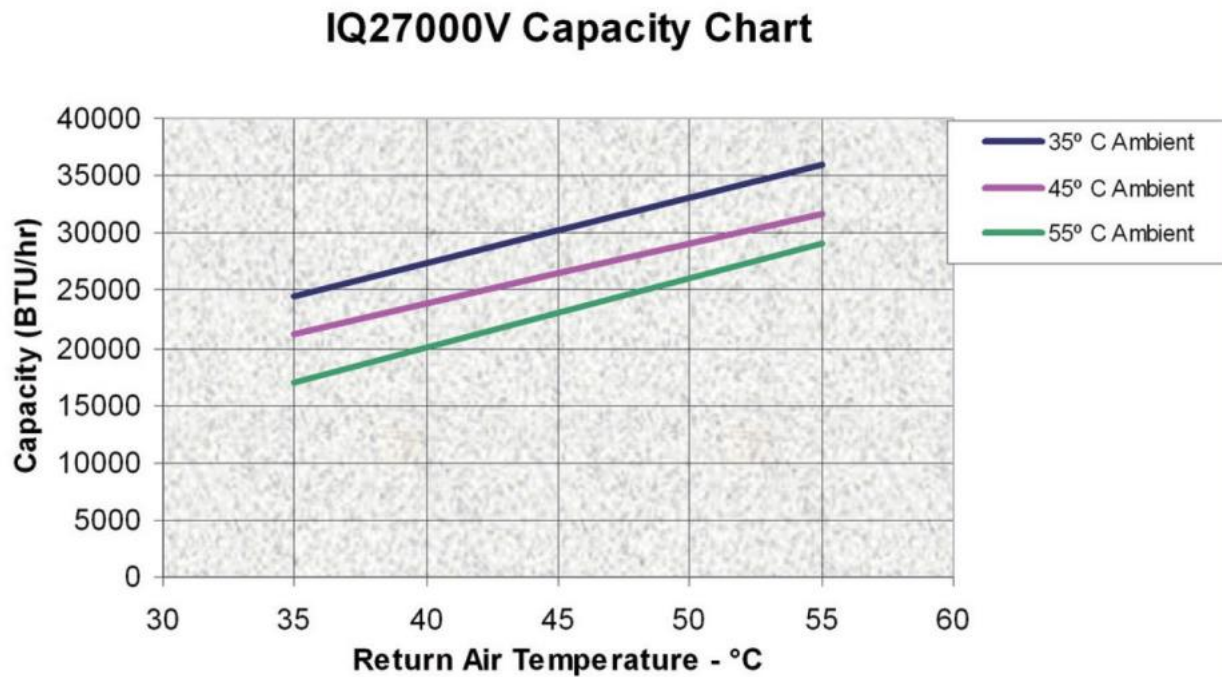


Figure 3.23 Ice Qube cooling capacity chart.

Ice Qube cooling capacity over return air temperature for each ambient temperature at the condenser location. At worst external conditions, 55°C (131°F) ambient at 35°C (95°F) return air temperature, the cooling capacity is rated for 17200 BTU/HR. At requirement conditions for RECON, 115°F ambient and 95°F return air temperature, the system is rated for 21500 BTU/HR.

3.6 Framing Material Selection

3.6.1 Approach

Given the selection of the HPC computing stack, ruggedized case system, cooling system, the framing and ducting design must integrate the systems together to fulfill overall functionality. Framing and ducting design must ruggedize the overall system for transport and operation in edge environments, provide closed circuit cooling resistant to dust and water ingress, and support ease of operation required. System framing must be substantial enough to

support the weight of either the computing stack or air conditioner systems or both, however, not add excessive weight that would impede the mobility of the system without lifting equipment.

3.6.2 Framing Hardware Selection

80/20's 15 series extruded aluminum profiles were selected as the primary structural component series for RECON's prototype effort. 80/20 T-slot profiles provide lightweight, corrosion resistant extrusion framing members and supporting hardware, which provide multifaceted approaches to mechanical framing solutions. Material tensile strength is 35000 PSI, made of 6105-T5 grade aluminum, with a clear anodized finish. Assembly of these profiles for the structure allows for modular design compared to welded framing. T-slot surfaces on 80/20 extruded profiles accept hardware that can be used to mount joining brackets, panels and other aluminum extrusions, structural members, seals and gaskets. 80/20 extrusions provide high feature density allowing for flexible use of the members with minimal machining costs. Lightweight extrusion profiles provide additional light weight design flexibility. 1.5" by 1.5" square and 1.5" by 3" profiles are primarily used for framing design, alongside supporting 1.5" x 1.5" aluminum angle extrusion.

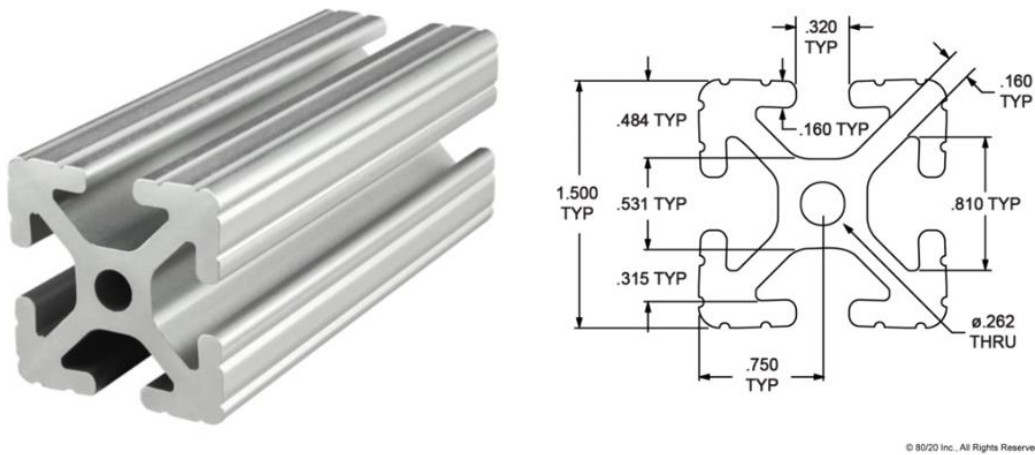


Figure 3.24 80/20 1515 series extrusion profile and dimensions.

1515 extrusion with cross section dimensions. Center hole can be tapped to accept 5/16"-18 fasteners, and fastening hardware is inserted through t-slot sections on each face. 1515 section weighs .1123 lbs. per inch.

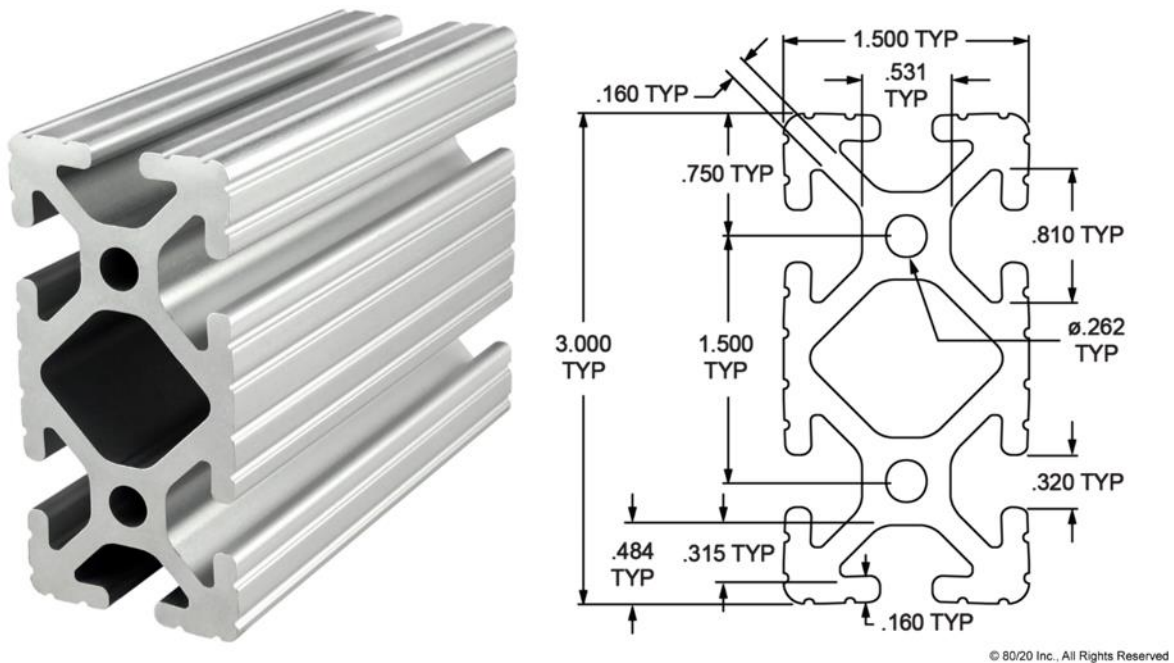


Figure 3.25 80/20 1530 series extrusion profile and dimensions.

1530 extrusion weighs .2025 lbs. per inch.

Joining methods used in the 80/20 15 series members yield high bending moment resistance. 1515 series and 1530 series members joint by t-slot end fasteners, where t-slot inserts hold 5/16"-18, 3/4" or 1" threaded length fasteners which connect perpendicular 1515 or 1530 members flush to the edge face, yield 410 ft-lbs. and 820 ft-lbs. cantilever moment load until failure. External surface brackets can be used to connect members singularly or in reinforcement, providing additional 200 ft-lbs. of cantilevered moment resistance until failure.

Table 3.1 80/20 15 Series t-slot extrusion properties.

Extrusion Part	1515	1530	1515-Lite	1530-Lite
Weight, lbs. per inch	.1123	.2025	.0873	.1679
Maximum Moment Inertia in.^4	.2542	1.8042	1.3847	.1853
Moment capacity: End fastener, ft-lbs.	410	820	250	500
Moment capacity: Joining bracket ft-lbs.	100	200	100	200

Common extrusion properties are 6105-T5 grade with 35000 PSI yield strength, moment capacities are tested from cantilevered loads applied 6" from the connection point.

3.6.3 Duct Material Selection

Similar to the computing stack's Impact case, 5052 aluminum alloy was selected for body and ducting panels. Lightweight, weldable and relatively formable compared, 0.09" thickness 5052 alloy aluminum sheet metal was selected to construct ducting and body panels for the anticipated 24" diameter average panel size or ducting faces on prototypes.

3.7 Concept Design

Given the selection of the full RECON computing stack, Impact ruggedized server rack case, the Ice Qube IQ27000V condenser and evaporator cooling system, and 15 series 80/20 framing focused framing solutions with aluminum ducting approach, a concept design was created. This design focused on establishing a layout of the components which integrated the systems together for a compact and ruggedized cooling system adapted to the Impact case. The evaporator air handling unit is positioned below the computing case so that inlet and outlet cutouts on the lids correspond to intake and exhaust ports, aligning the direction of cooling air circulation into a closed loop circuit. 80/20 framing systems are applied to secure the evaporator in position, mount the computing case above, provide mounting geometry for the required ducting, and enclose the ducting and AC systems from side impacts. The condenser is mounted on extended framing behind the cooling circuit, dumping waste heat backwards.

This design minimized the overall dimensions of the framing system, and overall length of the unit, focusing on a modular and compact layout. The weight and strength of the framing was required to support the total weight of the computing case and cooling systems, requiring additional vertical supports and moment braces. The addition of electrical control panels and supporting equipment were anticipated and to be refined in further development. It was determined that this concept layout could successfully fulfill remaining RECON requirements with further development.

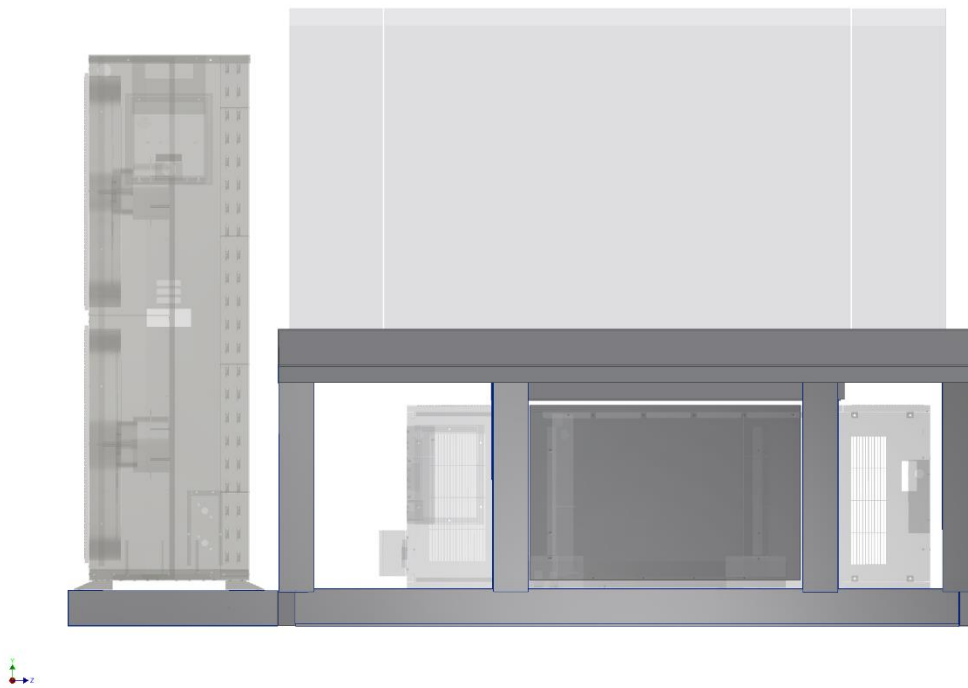


Figure 3.26 RECON integrated component concept design side view.

1530, 1515 and angle extrusion are arranged to create an integrated platform and housing for the AC components and ruggedized computing case. The right-side feeds supply air into the case from below, the return air is pulled through the larger vents left of center.

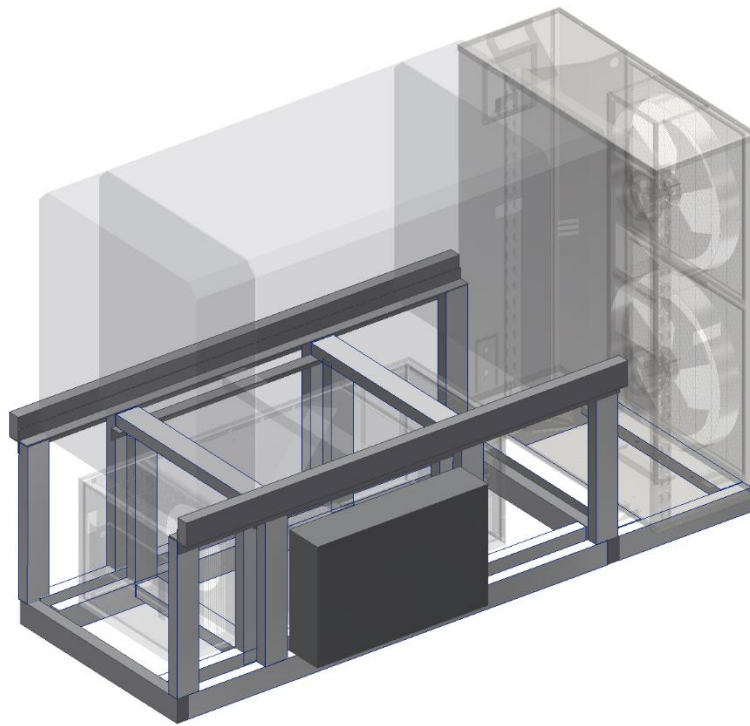


Figure 3.27 RECON integrated component concept design isometric view.

Framing is arranged to allow the computing case to rest above evaporator profile without additional overhead framing. Internal framing secures the AC and anticipated ducting design, while the external framing layer transfers load from computing case down to the base layer. Caster wheels can be added below the frame

CHAPTER IV

PROTOTYPE 1 DESIGN AND EVALUATION

4.1 Overview

The selected individual systems are integrated into a combined platform to fulfil the requirements of RECON. Two iterations of RECON (Prototype 1 and RECON) are designed and constructed in response to design objectives and priorities, this chapter focuses on Prototype 1. Prototype iterations are evaluated to determine functionality, and to adjust design parameters for the next iterations, influencing future considerations.

4.2 Design Objectives

Given the selection of the computing components and their ruggedization, the cooling system, and the general approach to system framing and material choice, the remaining design integrates individual functions and details of operation in order to fulfill objectives and requirements. Objectives such as ruggedization, compact design and transportability were additionally prioritized for iteration 1 design over the base requirements. The overall approach was evaluated and relating performance metrics established following each iteration, leading to reprioritization of objectives and refined outlooks on future design.

Prototype 1 Attributes and Objectives

MOBILITY	SURVIVABILITY	MANUEVERABILITY
<ul style="list-style-type: none"> • Transportability • Minimal system dimensions • Minimal weight • Modular components 	<ul style="list-style-type: none"> • Ruggedization, shock and impact resistance • Resilient structure and framing • Environmental protection and insulation • Cooling capacity against high ambient temperature • Ingress protection against dust and water 	<ul style="list-style-type: none"> • Ease of operation and deployment • Access to controls • Transportability during operation • Flexible power management

Figure 4.1 Prototype 1 attributes and objectives.

Prototype 1 prioritized objectives are bolded over standard objectives.

4.3 Design Response Overview

Given the selected components, the remaining system design was focused on fulfilling the remaining requirements. It was imperative that the system integration and framing was able to deliver the computing systems to edge environments, creating a unique HPC resource capable in edge conditions. Prioritization was placed on the corresponding aspects of mobility and survivability necessary to get the system to those environments, influencing design to prioritize transportability, compact in design and transport, and additionally ruggedized framing.

The general orientation of main systems was carried over to prototype 1 design, placing the ruggedized case above the cooling system, mounted on framing which ducts and mates the computing case into a closed-circuit cooling loop, which can be lifted onto and off the framing. The condenser orientation was rotated 90° in order to better fit through facility doorways.

Necessary hardware and controls features were added to achieve remaining requirements of mobility and operability. The overall dimensions are 105" length, 31.25" width, and 60" in height. The case and evaporator dimensions are 66" length, 31.25" width, and 60" in height. The detachable condenser unit and framing dimensions are 48.5" length, 29" width, and 55.25" in height.



Figure 4.2 Prototype 1 design render isometric view.

Ruggedized case passively mated to closed circuit cooling system below, condenser mounted on detachable framing, behind.

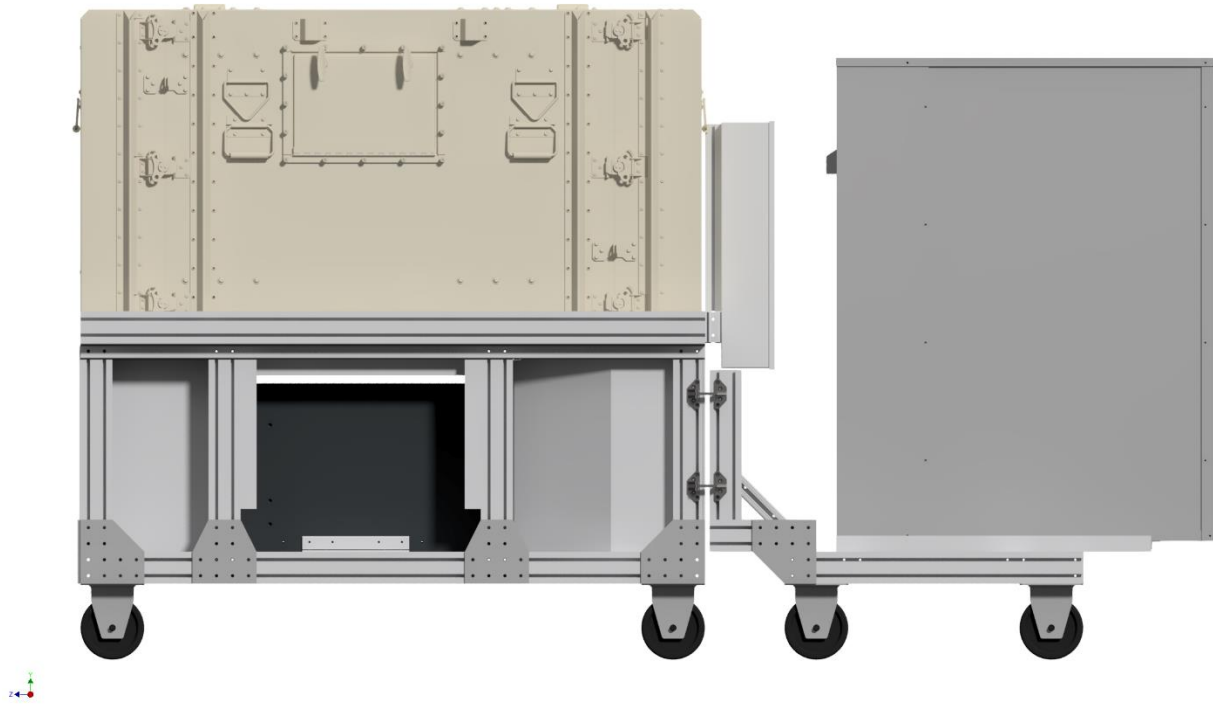


Figure 4.3 Prototype 1 design render side view.

Electrical breaker panel and distribution box was mounted behind the ruggedized case, in line with the condenser to minimize system width, compared to concept.

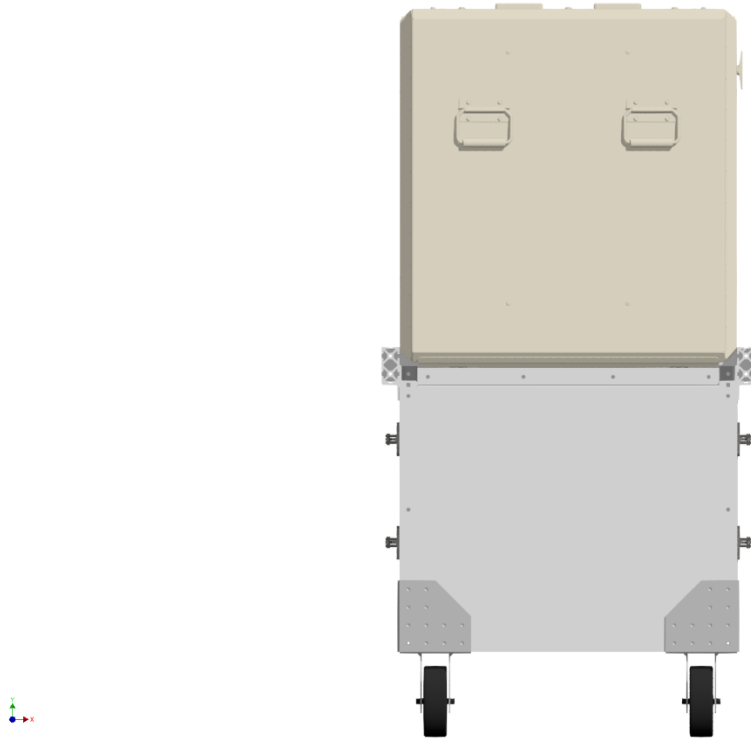


Figure 4.4 Prototype 1 design render front view.

Widest features are the framing rails, which laterally contain the ruggedized case and prevent side impacts from affecting hinge hardware during transport.

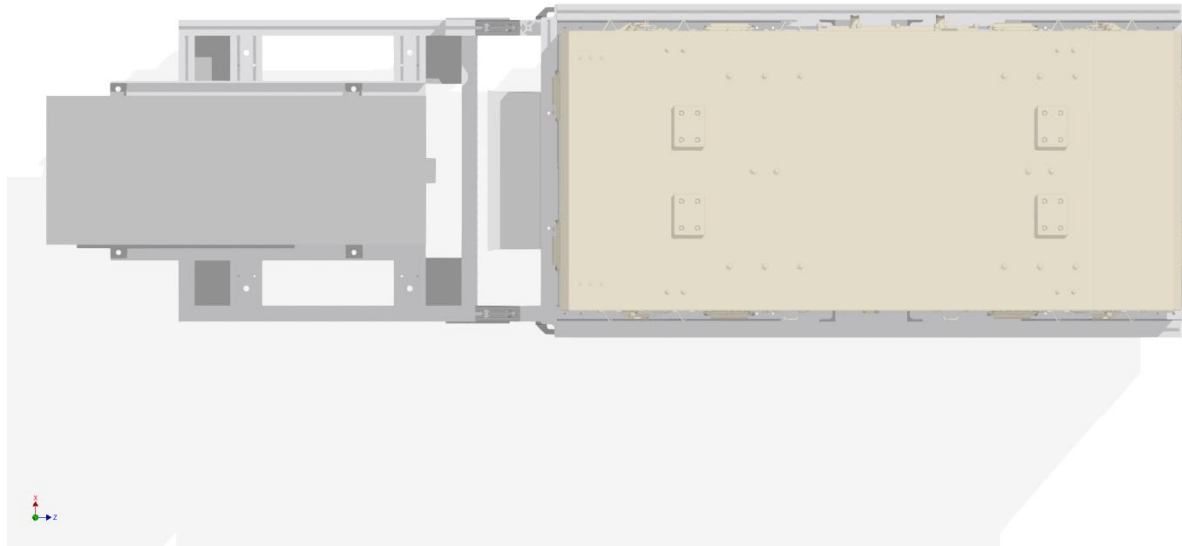


Figure 4.5 Prototype 1 design render top view.

4.3.1 Framing Hardware Design

1530 and 1515 80/20 extrusions primarily compose the structure, using various t-slot prefabricated joining methods available. The framing system integrates the computing and cooling components into a transportable and ruggedized cohesive system, housing and securing AC system and ducting, the ruggedized case and required interfaces. Transportation of equipment to edge environments was expected to subject them to shock and loads likely more severe than on site operation, influencing the design to use redundant fastening, multiple layers of framing and heavier extrusion profiles.

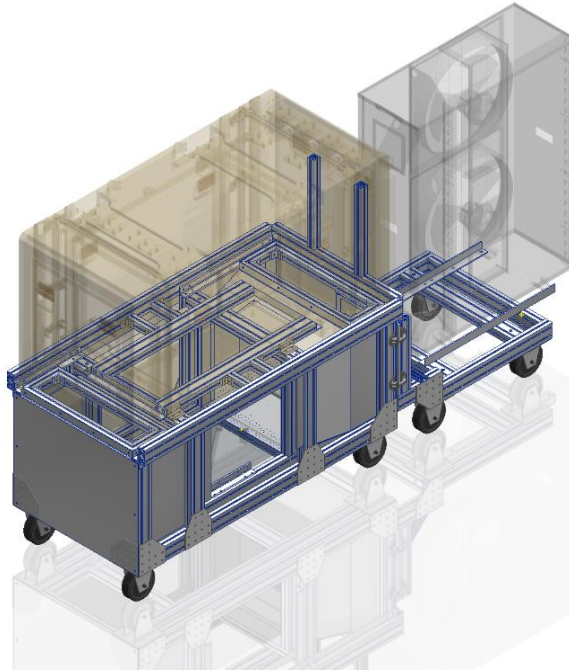


Figure 4.6 Prototype 1 framing design render isometric view.

The framing design approaches its functionality by four sections: The lower base chassis section, vertical supports and evaporator housing, ruggedized case interface and condenser framing.

4.3.1.1 Base Chassis and Hardware

The lower level of framing forms a chassis using four 1530 extrusion members across the whole length of the case and evaporator section, capped by additional 1530 members which cap the end across the width. Each corner mounts a caster wheel assembly with individual load capacities of 300 lbs., critically enabling ease of transportation to the entire prototype. The inner rails primarily support the evaporator housing, and the outside rails support the vertical members

that support the ruggedized computing case. The same separation is maintained in order to allow for impacts sustained across the lower sides to not be directly transferred into the evaporator and ducting systems. The long channels left along the edges additionally allow for larger wheels to be installed without adding to the overall height of the assembly. As the base for the vertical members, additional joining bracket are placed outside each perpendicular connection, additionally stiffening the system against lateral loads.

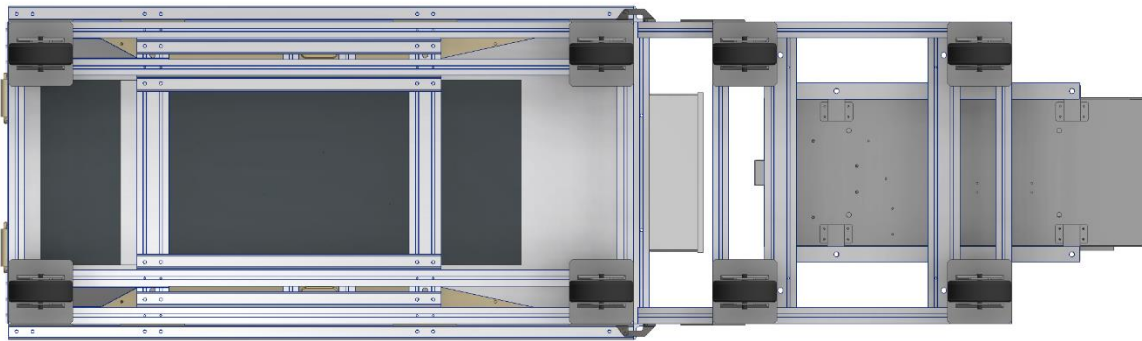


Figure 4.7 Prototype 1 design render framing and hardware bottom view.

4.3.1.2 Vertical Framing

The vertical members are primarily made of 1530 members. Twelve vertical members connect the lower chassis to the upper ruggedized case interface, eight of which have additional joining plate stiffening in addition to the typical end connector joining method. 1515 standard end fasteners add 410 ft-lbs. of cantilever moment resistance per connection, at least doubling per 1530 connection. Joining plates add 200 ft-lbs. per perpendicular joint. Total moment resistance offered across the eight outside vertical supports provide 8160 ft-lbs. of moment resistance, providing stability for the mounted computing case against translational body impact during transportation. Four internal verticals are placed to secure the evaporator and ducting panel assemblies, also securing angle bracket sections part of ducting and sealing airflow.

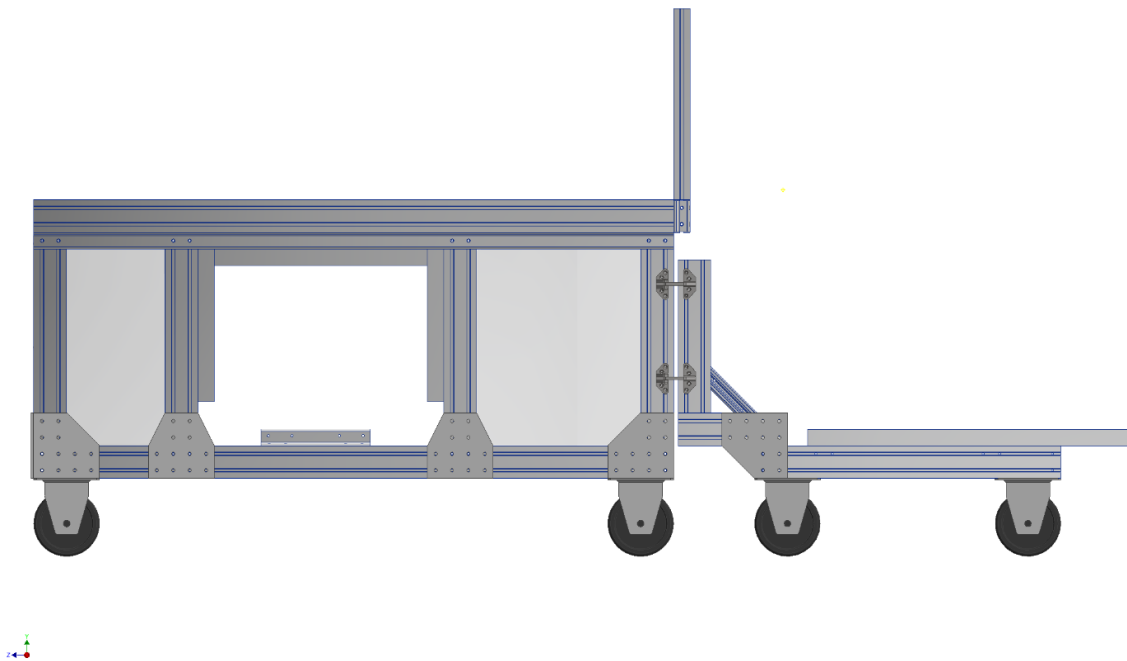


Figure 4.8 Framing design render side view.

Vertical members connect to upper-level offset rails via angle profiles running partial lengths of the assembly.

4.3.1.3 Upper-Level Case and Duct Framing

Upper-level framing secures the Impact case and ducting connection features. The upper-level framing geometry conforms to polymer bumpers and protrusions underneath the Impact case, and the outside dimensions of the case. Those features support Impact case stacking and are capable of being used as mating features and are used to secure the case within the confines of the frame. Part of a modular design approach, the mating features guide the case into position when being lowered onto the frame, ensuring correct alignment of the case over the ducting features, ensuring the seals are pressed to complete closed circuit airflow. No direct fastening or connections are required for operation. The outside rails prevent the case from displacing horizontally and the internal framing secure the case from sliding forwards or backwards.

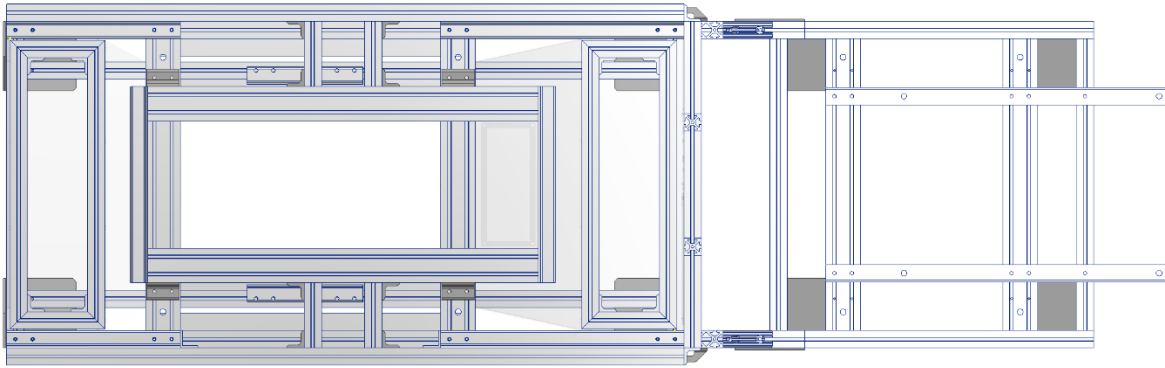


Figure 4.9 Framing design upper level render, top view.

Assembly of framing is shaped to secure the Impact computing case without fastening connectors, also sealing the closed-circuit ducting inlets and outlets.

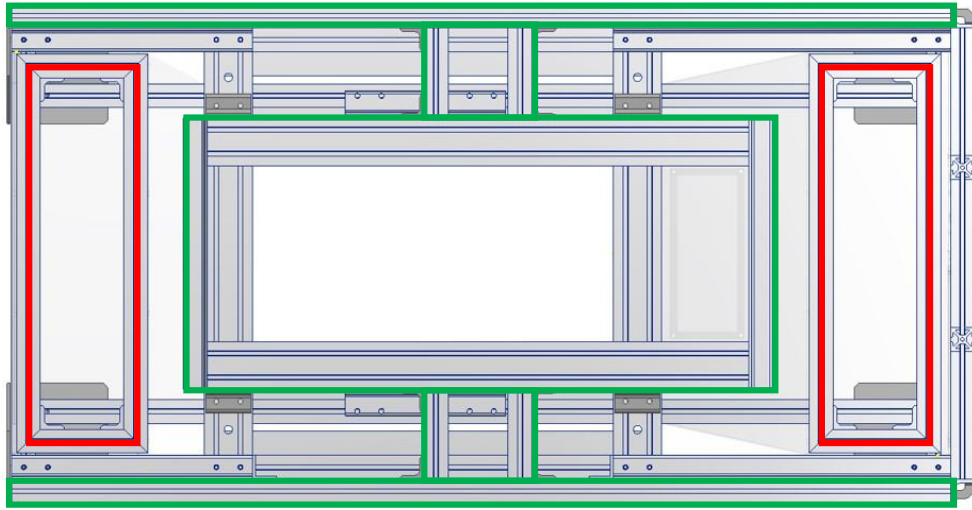


Figure 4.10 Case mounting and duct position framing design render, top view.

Case mounting geometry is highlighted in green, and duct sealing geometry is highlighted in red.

4.3.1.4 Condenser Framing

The condenser was mounted on its own framing sub-assembly to increase ease of operation and transportation. The framing assembly can hinge or detach from the main body, allowing the entire unit to maneuver through facilities better, decreasing the required width of hallways and turns within facilities needed for this prototype to fit through. Additionally, the detachable condenser allows waste heat to be dumped outside the operating environment, for noise and waste heat separation. Flexible refrigerant lines are installed in order to provide the necessary range of motion. Condenser mounting tabs are secured to the frame with dampening rubber washers for additional vibration and noise reduction. Four caster wheels enable system mobility.

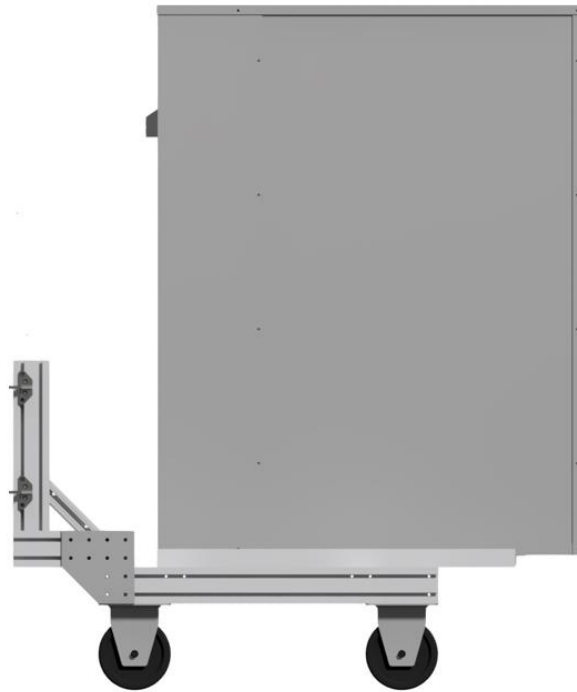


Figure 4.11 Prototype 1 condenser sub-assembly framing design render, side view.

Condenser unit mounted on separate framing system. Hinge connections allow for hinged or detached operation. Machine weight is biased to the left, stability remains when detached.

4.3.2 Cooling and Ducting Design

The core functionality of cooling the computing component during operation is achieved by securing and ducting the IQ27000V airflow into the computing case, maintaining ingress resistant closed-circuit airflow. The evaporator unit is integrated into the framing so that welded ducting structure can surround and seal the airflow, directing it into and out of the case above. The condenser is connected to the cooling system using flexible insulated refrigerant lines, while integrated for that functionality, the evaporator unit can still be removed for replacement or maintenance, by removal of front and back ducting panels which attach only to the framing system by fasteners.

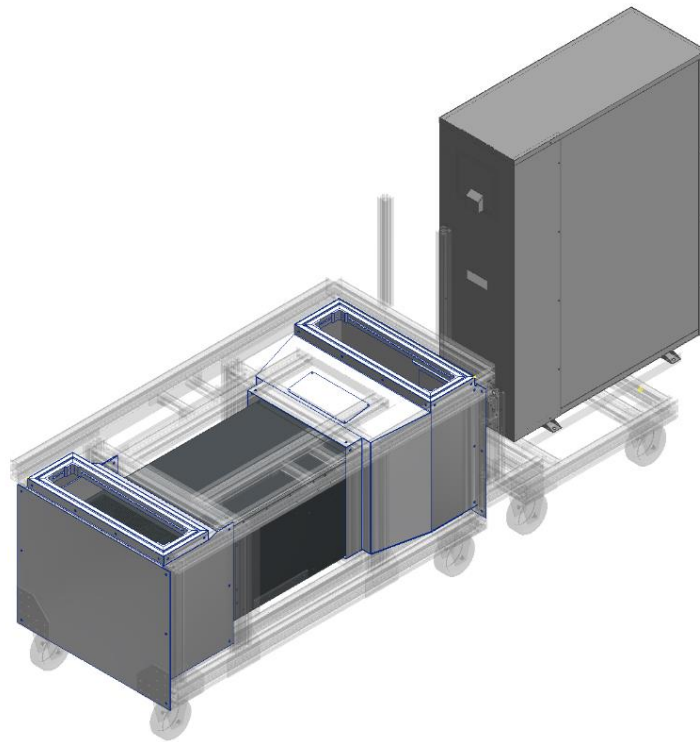


Figure 4.12 Prototype 1 ducting and cooling design render, isometric view.

Isometric view of cooling and ducting specific features. Ducting adapts the airflow from the evaporator unit into the computing case.

Ducting sections are constructed of .09" thickness 5052 Aluminum. Two sections of duct panels are welded together to direct the intake and exhaust flows of the computing components, featuring sloped walls which attempt to aid in efficient redirection of the flow out of the evaporator vents. The 80/20 framing housing the gasket mating surface to the case is also mounted directly to the duct structure. Insulation boards and additional panels around the evaporator body are added to reduce convective and conductive heat loss through the system.

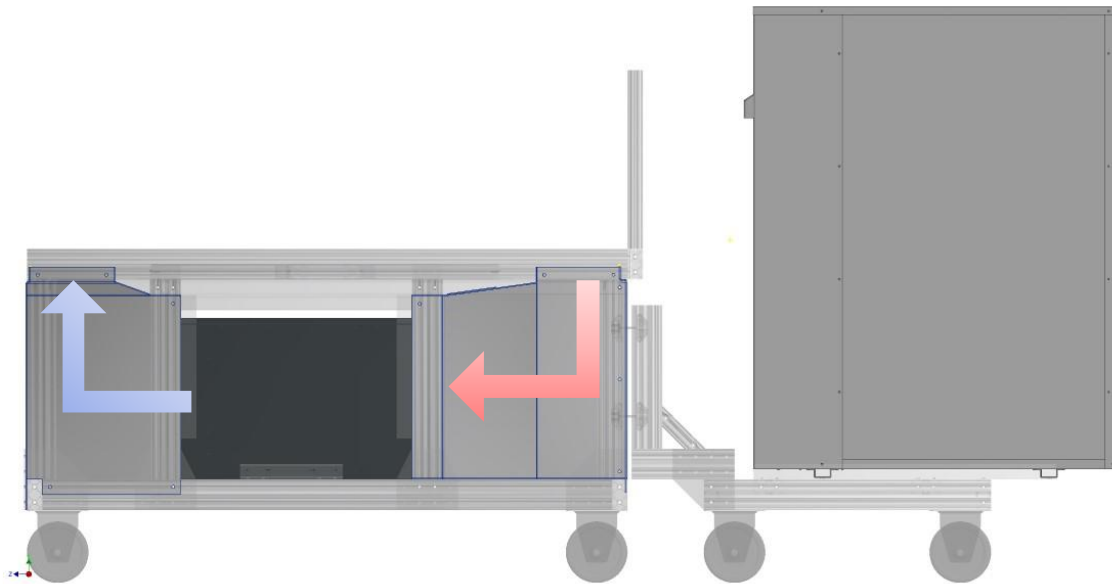


Figure 4.13 Prototype 1 cooling design closed circuit airflow.

Side view of IQ27000V system, airflow direction and cooling visualized across evaporator heat exchanger in center.

4.3.3 Electrical and Control Systems

Electrical and AC control interfaces are located on the rear of the main body. The AC system uses 240V, while the computing components use 120V, either supplied via generator or shore power. A breaker box is configured to control amperage to the Evaporator and condenser, and another breaker to the computing case. Two 30 Amp slow burn joined breakers supply the AC components, and the two 20 Amp breakers protect the computing components. 120V periphery devices are controlled over two 20 Amp breakers. Electrical lines and power from the breaker box are directed to the bulkhead connector panel on the Impact case, the evaporator, the condenser, and an additional 120V socket outdoor receptacle supplying power for peripheral equipment, such as monitors and control computers. Attached electrical cabling, distributed from

the electrical breaker box, include the 14-50P plug for main power that interacts with 230-250V 50A 14-50R receptacles. A L6-30R cable power cable carries power to the Impact case, powering the computing stack.

For remote operation lacking facility shore power, a generator is specified for the expected power usage of 13000 Watts, generating 240V as well as 120V. While not directly part of the RECON system, the system will be expected to operate with generators on site. The generator is separately transportable. A DuroMax power systems XP13000EH generator, capable of running on propane and gasoline, is used for the design and testing of RECON systems.

The control system native to the IQ27000V is used, which has configurable cooling set points and thresholds. This control interface is detached from the evaporator main body and mounted on the rear ducting panel.

4.4 Prototype 1 Evaluation

4.4.1 Overview

The construction of Prototype 1 of RECON is reviewed. Design changes and notable features during construction are discussed leading up to the as-built results. Prototype 1 is evaluated for its performance of objectives, establishing the current system performance for comparison. The prototype's performance is considered for further iteration of objectives and system design.

4.4.2 Construction

Two framing units and one full prototype and cooling system were built for the first design iteration. A Nvidia DGX-1 primary computing stack was built for this iteration. Component placement within the Impact ruggedized case required slight modification to the rack

structure, where material needed to be cut out to accommodate its main body and mounting position. Rack strength was not impacted. The DGX-1 and DDN hard drive were placed lowest on the rack, minimizing the center of gravity and placing the most heat intensive units closest to the inlet and exhaust ports. Communication and power cables were routed between the computing stack and to the bulkhead communications access.



Figure 4.14 Prototype 1 ruggedized computing stack.

Nvidia DGX-1 computing pipeline. Component intake is above cool air supply.

Two framing systems were constructed using the selected materials. Assembly of the framing levels from lower chassis, proceeding up through the vertical duct housing, evaporator

and case supports extended slight angle tolerances in the cuts forming each member from stock extrusion. Rail positions and duct framing needed to be adjusted to meet tolerances for the duct assemblies, which already experienced warping due to welding.

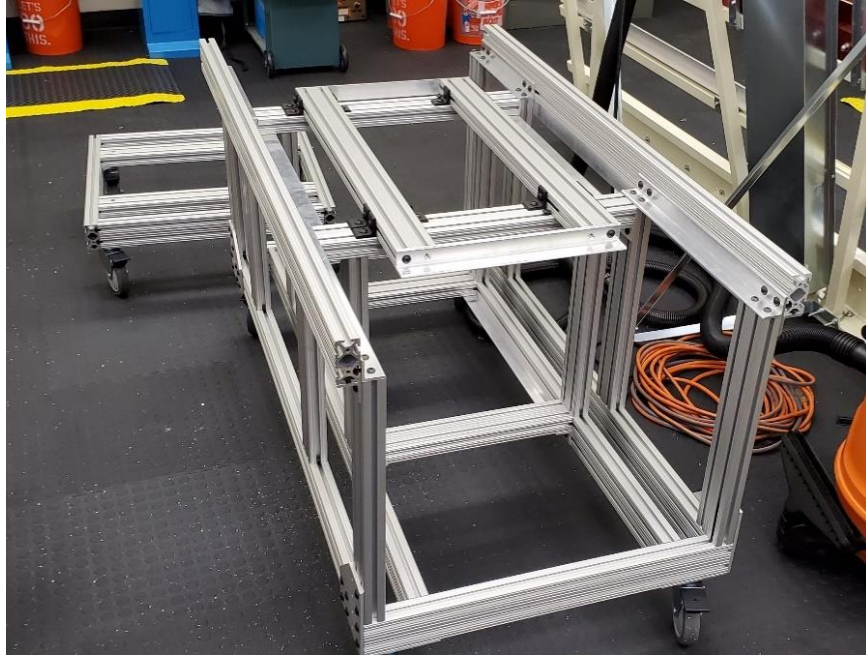


Figure 4.15 Framing construction Prototyping 1.

Ducting assemblies were installed following framing assembly. Evaporator placement followed duct and sealing placement. Rubber strips seal above and below the evaporator body and are clamped down by the fastening of upper cross support rails, in addition to angle brackets on the sides of the evaporator body directly fastening to the structure. Additional Tube gaskets mounted on angle extrusions were applied to seal the sides from ingress, securing closed circuit airflow. Silicon sealer and foam was used to seal seams and t-slot air leakage paths before the installation of foam board insulation across internal duct faces. R-3 and R-5 rated insulation was

added to exposed evaporator body faces to reduce heat gain. Direct modification to the cooling components were limited in order to maintain ease of component replacement and maintenance. Modifications made to the evaporator include the repositioning of vents to face upward as needed for this prototype's orientation, and the control panel was separated from the evaporator body and remounted to the rear duct panel. The rear duct panel was modified as needed to mount the controls and pass through refrigerant and power lines.



Figure 4.16 Evaporator and partial ducting with framing.



Figure 4.17 Construction framing and partial ducting without side rails.

Peripheral changes to the framing and ducting accommodated the details of electrical component installation and hardware interaction. Additional framing members were added to introduce cable management for power and refrigerant lines. Cutouts were added on the upper side rails in order to provide access to lid latches. Condensate lines were added in order to dump water externally.

4.4.3 Prototype 1 As-Built Results

The following photos show the functional RECON prototype 1, with the Nvidia DGX-1 computing stack. The condenser is separated and operating as shown.

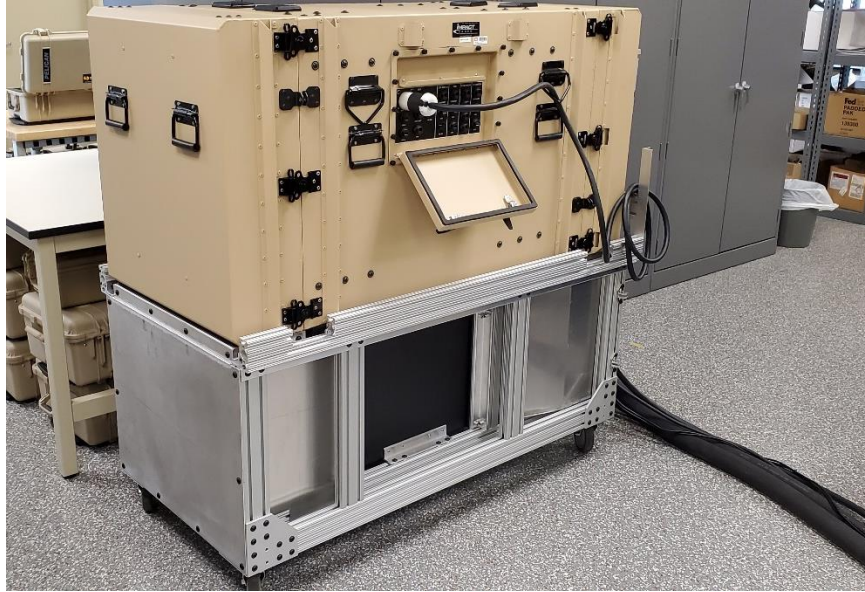


Figure 4.18 RECON Prototype 1, main body computing case, evaporator framing and ducting.



Figure 4.19 RECON Prototype 1 side view, evaporator main body.

Additional side panel insulation of exposed evaporator body was not yet added.



Figure 4.20 Prototype 1 as-built main body rear view.

Rear of prototype 1 shows the refrigerant, power and condensate line pass through the rear ducting panel. The control panel is mounter below, and the electrical distribution and breaker box, above.



Figure 4.21 Prototype 1 condenser system as-built.

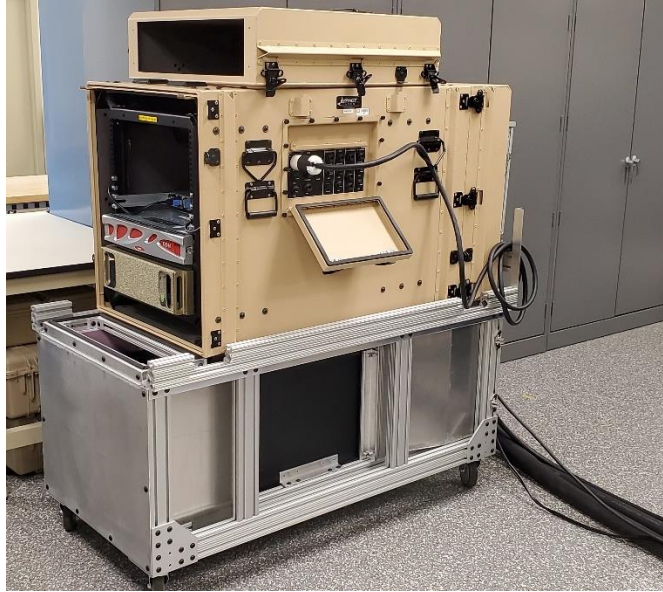


Figure 4.22 Prototype 1 main body with detached case lid.

4.5 Evaluation

The RECON prototype was evaluated over a series of demonstrations and tests of individual functionalities. Although this iteration was not tested for its maximum processing capability in 115°F 100% humidity weather on shore power, the surrounding functionalities were tested enough to evaluate if the design fulfilled requirements, establishing baseline capabilities.

4.5.1 Component Functionality

With the initial construction of the computing stack and the modular cooling system complete, the electrical and AC systems were tested for their functionality. The AC system flexible lines were connected and charged, and the AC control system was set to activate cooling at thresholds below the room temperature, 73°F, of the facility. Using the facilities 240V power, the AC was able to produce a constant cool air stream at least 10°F below the ambient air intake drawing through the case, with one lid removed, similar to Figure 4.22. Following open

operation, the DGX-1 stack was powered on, the front lid was attached to close the airflow circuit, and the cooling system was able to rapidly cool the system loop within a minute, down to 63° before the control powered down the condenser, where more 10 minutes passed before the internal airflow temperature reached the kick-on point. A low to mid-level processing load, relative to processing capability, was run on the DGX-1 within the same environment. Kick on and kickoff times of the condenser began to equalize, however, the cooling cycle was much shorter than the time it took to raise the system temp back to the cooling threshold, indicating that the system can cool with ease in the initial test's moderate environment.

4.5.2 System Mobility

Throughout the construction and development of this prototype, the entire system was transported within facility and between different facilities, through either accessible or entrances with a few partitions or steps, showing both ease and limitations in its maneuverability and transportability. Once the computing stack was installed into the ruggedized case, the Impact case and computing equipment weight neared 400 lbs. Ideally the computing case would be separated from the cooling system during transportation, requiring that at each transportation step, the case would need to be lifted from the cooling system, set down, opened in order to retrieve stowed caster wheels, and again lifted in order to place the caster wheels and reinstall the duct cutouts on each lid. The process required four people, without lifting equipment, to handle system deployment. The manipulation requirements of this configuration limited ease of operation, ultimately reducing maneuverability despite the compact form factor.

4.5.3 Maneuverability

Once deployed, the prototype maneuvered well through facilities. The 31.25” width fit most main facility doors and paths, and the hinged and detachable condenser framing allowed for navigation through doors facing hallway widths at least 64”. The system could be easily moved through flat plane environments, passing bumps, doorjams and seals, slab breaks, etc. successfully, given the obstacle height did not exceed the wheel radius of 2.5”. The entire functional unit could be loaded into trailers, and when separated, could be transported on pickup truck beds.

4.5.4 Survivability

Combined component systems contributed to the ruggedization of the computing stack and protection from the environment during operation and transportation.

4.5.4.1 Ruggedization

Survivability requirements were met due to the continued operation and function of computing components, frame integrity, and environmental protection throughout the transportation and operation of the prototype. The continued function of all components through cumulative exposure to shock as needed for the use case was counted as requirement fulfillment. The computing components were not subjected to deliberate shock until failure, however, the full prototype as well as alternate computing stacks in identical Impact cases were transported across the facilities, exposed to ground transportation and shipping shock and vibration. Similar computing stacks also mounted in identical Impact cases to RECON’s computing stack, housing the Dell 7920 Precision and supporting components, were shipped via ground transportation across the United States. All units arrived fully operational, fulfilling the ruggedizations

requirement against shock and impact subjected by transportation. Maneuvering the deployed prototype often subjected the system to glancing and direct impacts into wall corners, doors and door jambs, other equipment, vending machines etc. leaving the system unaffected. A pointed front impact would likely damage the front ducting panel, however, so additional design features could ensure full framing resiliency, if the dimensional minimalism was relaxed. Overall, the framing and computing systems were resistant to any encountered shock and vibration.

4.5.4.2 Environmental Protection

Survivability objectives were achieved across the instance that the prototype was operated outside. The prototype was operated outside in humid, 80°F - 85°F temperatures with 50-70% relative humidity with low to medium load processing functions running on the DGX-1. AC systems functioned similarly to the indoor setup tests, with a relatively short cooling period, where the low temperature threshold was reached, and the condenser cycled off. Overall temperatures according to the AC control unit cycled from 63°F to 78°F, depending on settings. Computing components did not throw any humidity or temperature alarms during operation. No dust or signs of condensation were found within the case after outside operation. Nominal environmental survivability was achieved in moderate conditions.

4.6 Results

Operation of Prototype 1 proved basic functionalities were achieved for RECON requirements. While attribute and objective priorities initially emphasized for this design were accomplished, a rebalancing of priorities is required for the next iteration, to better match and perform in expected environments. Further testing is required in order to fully test the capabilities and limitations of the cooling system.

4.6.1 System Ruggedization

Ruggedization for transportation of the system was achieved, providing the shock and vibration resistance required for transportation on ground vehicles, including the framing and ducting. Focus on transportability by minimizing dimensions can be reduced, in order to provide additional impact protection to the ducting panels. Ruggedization of the framing provided redundant protections against impacts at the cost of system weight, however the base survivability objectives were accomplished.

4.6.2 Environmental Protection

Environmental protection of components and ruggedization during operation was achieved during outdoor conditions, where moderate outdoor conditions were applied to the system. The design maintained an IP64 capable resistance to dust and water, providing insulated airflow to the computing components. The system was not tested to full capacity, however, with high ambient temperatures with full processing load, the 17200 BTU/HR cooling capacity at 115°F ambient air temperature with 95°F internal return temperature remained unconfirmed. The cooling system needs to be monitored at setup to ensure refrigerant pressure conditions are optimal and leaks are not present.

4.6.3 Modular Operation

The modular design requirement was accomplished, in addition to the compact design, maximized the transportability of the system when separated or deployed. The configuration of modular components made the process of deploying the system difficult, where four people capable of carrying at least 100 lbs. off center were required to mount the computing case on the cooling system, limiting ease of operation, which limited to maneuverability.

4.6.4 Shore Power Operation

Operation of the system on shore power was accomplished. System functionality on generator power was successful, however, similar to the cooling system, was not operated during maximum capacity cooling and full processing load. The breaker panel successfully manages overcurrent, for example, on condenser over draw, and the Raritan PDU logs power consumption and data, however, power quality is not managed internally.

4.7 Conclusion

Prototype 1 evaluation proves the functionality of this combination of modular systems achieve functionality according to RECON requirements, although limited to moderate environments. System ruggedization, transportability and modularity are confirmed to fulfill requirements, and system objectives and priorities can better be applied in further iterations for improved operation in expected edge environments. Further testing of the cooling, electrical, and computing performance is required for proper evaluation of RECON capabilities.

CHAPTER V

RECON DESIGN AND EVALUATION

5.1 Overview

Design of RECON is continued in response to Prototype 1 evaluation results. Core components remain the same, while design priorities are shifted, resulting in the construction of one fully operational RECON iteration. Components are added to expand RECON functionality, and the system is evaluated for edge capabilities.

5.2 Design Objectives

The design of the second iteration, RECON, continued design response to the evaluation of prototype 1. Priorities of compact overall dimensions, configurations optimized for compact transportation, and higher strength framing can be relaxed. Framing systems and ruggedization proved to be adequate in previous evaluations. It was determined that effectiveness of the system would be best improved by prioritizing objectives under maneuverability, assisting tor minimizing the manpower required for operation and deployment. Existing environmental protection should be maintained. The new priorities are shown in the RECON attributes and objectives visual.

RECON Attributes and Objectives

MOBILITY	SURVIVABILITY	MANUEVERABILITY
<ul style="list-style-type: none"> • Transportability • Minimal system dimensions • Minimal weight • Modular components 	<ul style="list-style-type: none"> • Ruggedization, shock and impact resistance • Resilient structure and framing • Environmental protection and insulation • Cooling capacity against high ambient temperature • Ingress protection against dust and water 	<ul style="list-style-type: none"> • Ease of operation and deployment • Access to controls • Transportability during operation • Flexible power management

Figure 5.1 RECON design attributes and objectives.

Objective priorities in response to evaluation of prototype 1 are bolded, where this second iteration design prioritizes maneuverability and ease of deployment.

5.3 Design Response

RECON design configuration was adjusted to the new priorities, while maintaining functionality across all requirements. The main components of the system, the computing stack and case, cooling system, and framing material remained the same. The configuration of those components was changed, and the framing and ducting design changed in order to accommodate the requirements. The new configuration mounts the evaporator unit above the Impact case, where articulated drop-down ducting can supply closed circuit airflow below to the computing case. This configuration does not require the case to be lifted for deployment or transportation, greatly increasing ease of operation and maneuverability. Additional framing features secure the computing case to the frame and AC, mount electrical controls, ducting and periphery devices.

The condenser is detachable, mounted on an independent frame, flexible refrigerant lines connect the cooling systems. The overall framing dimensional volume is increased, however, simplified ducting and framing design, along with lighter 80/20 profiles, aim to reduce the weight of the entire unit. Overall dimensions of RECON are 115" in length, 31.75" in width, and 60" in height. The main body is 74" by 31.75" by 60", and the condenser body while detached is 44.5" by 28" and 55.5" in height.

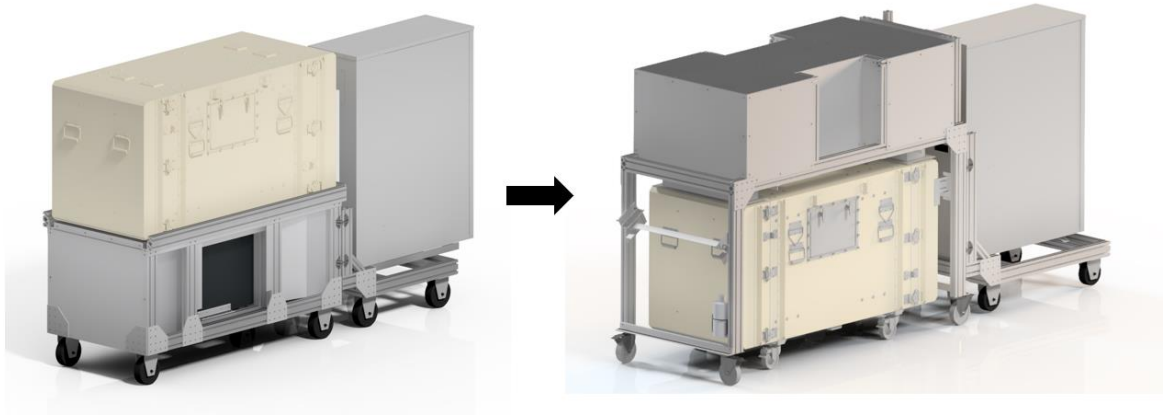


Figure 5.2 Prototype 1 to RECON design progression.

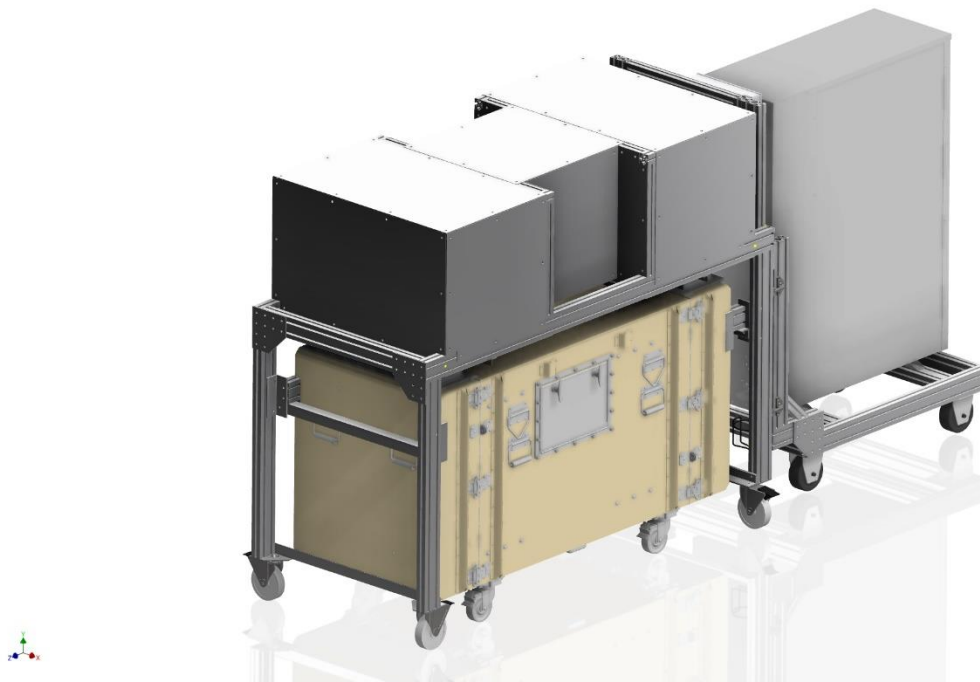


Figure 5.3 RECON design render isometric view.

RECON iteration with overhead cooling system configuration and detachable condenser. Computing case can be rolled away from AC cooling system and framing.

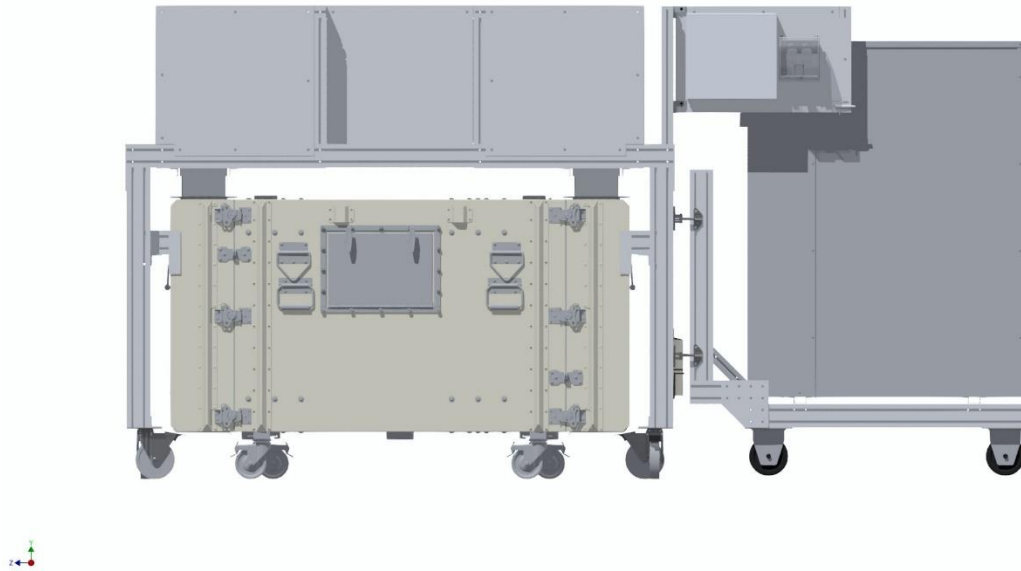


Figure 5.4 RECON design render side view.

Side view, drop down ducting and case lock in framing features are visible around computing case. Electrical control panel is hinge mounted.

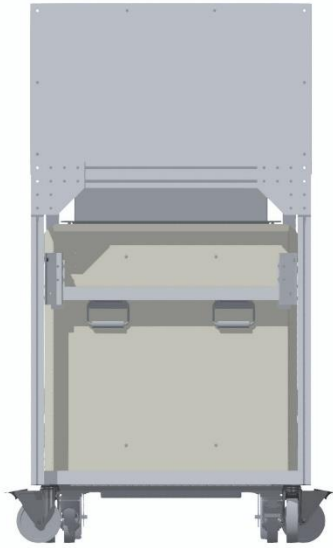


Figure 5.5 RECON design render front view.

The framing system provides mobility and encloses the front and rear of the system.

5.3.1 Framing

Framing design of RECON utilizes 1515-Lite and 1530-Lite framing members, compared to standard 1530 and 1515 extrusions. 1515-lite extrusions weight .0873 lbs. per inch, and weigh 22% less than 1515 standard extrusions, however, their moment of inertia is reduced by 27%. 1530-lite weighs .1679 lbs. per pound, a 17% less than the standard extrusion, at the cost of an average of 22% moment of inertia in both bending directions. The base strength of these extrusions is more than capable for their intended use and are used to inherently lightweight the system in addition to the design approaches. 80/20 15 series framing and joining hardware provide high strength relative to this application. With ruggedization priorities relaxed, as well as structural requirements, framing redundancy and vertical load supports can be minimized.

The framing supports the evaporator above the computing case, over legs that also support lock-in framing, which secures the case when mated to the cooling system. Integrated framing both holds ducting panels and secures the evaporator within the main tray. The condenser framing mounts the unit behind the main assembly, hosting the same detachable features as the previous iteration.

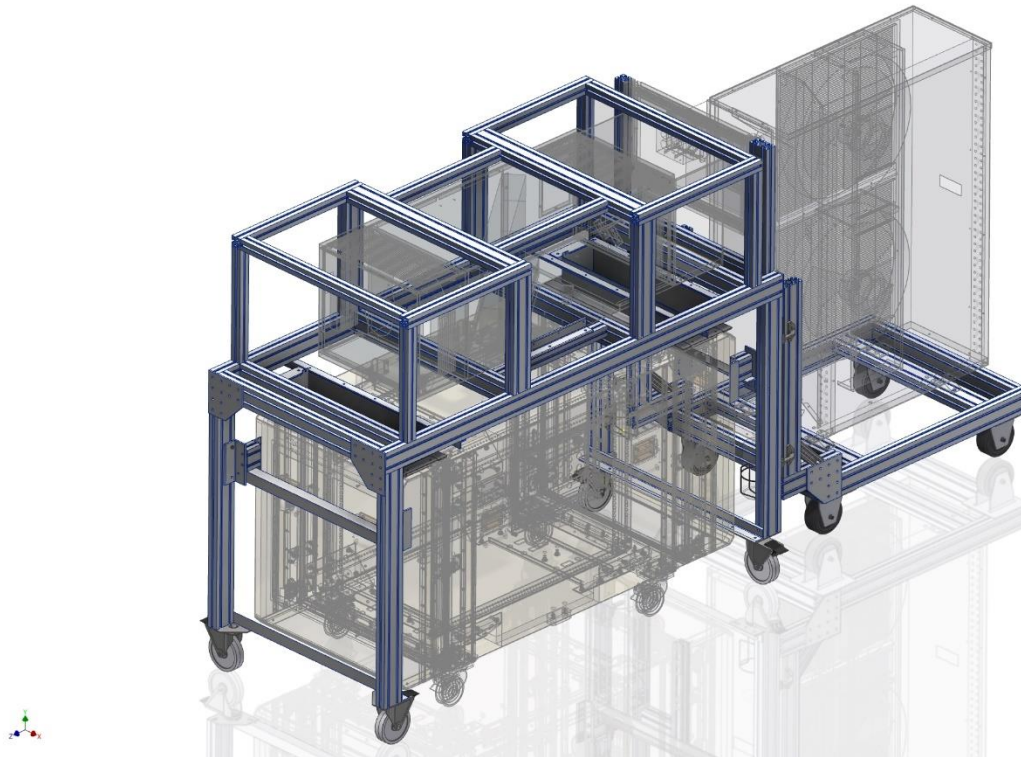


Figure 5.6 RECON framing isometric view.

Place all detailed caption, notes, reference, legend information, etc. Framing now only supports the weight of evaporator body, peripheral controls and devices, and ducting. Additional case lock-in features are added within the support legs.

Simplified framing design utilized only Lite extrusion profiles. The moment bracing of joining plates and end fasteners allows for full support of the overhead system using 4 points of connection. Two main side rails, with cross members corresponding to the ends, ducting panel

mounting points, and ducting intake and exhaust airflow separation lines, form the simplified main chassis or tray for the system. Overhead comparison of the design iterations shows the simplified and light weight structure, even with larger overall dimensions. The profile of the framing protects the ducting from forward impact, compared to side impact protection of the previous iteration, a trade off resulting from higher overall dimensions allowed for better maneuverability. Condenser framing is lengthened to place caster wheels further for a more stable stance.

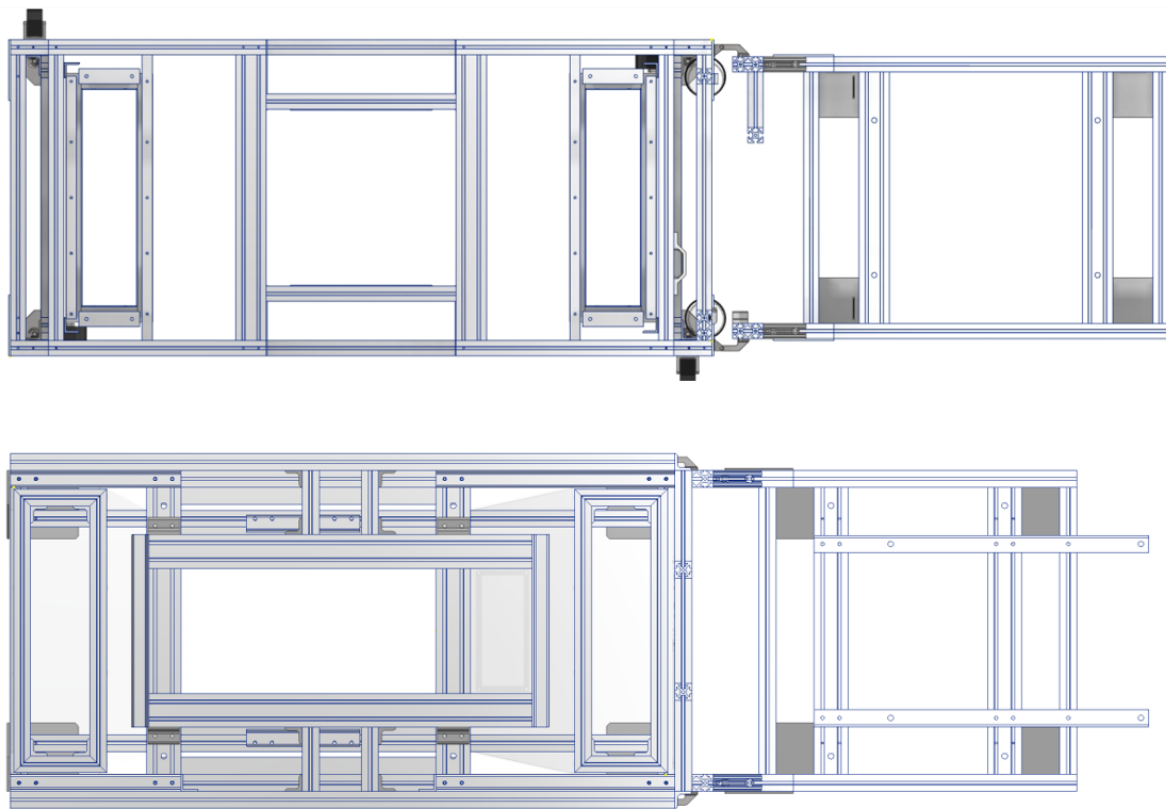


Figure 5.7 RECON framing top view comparison to Prototype 1 framing.

RECON framing is greatly simplified and features symmetrical parts for simplified construction and reduced weight.

Vertical framing and duct framing uses the 1515-Lite profiles primarily to hold flat duct panels along the outside edges of existing main rails and cross members. 1530-Lite double wide members are used to join adjacent panels along the body. The top frame above the verticals is designed to be built first under specific dimensions, so that vertical skew introduced by slight angle tolerances from perpendicular joining of vertical members can be averaged out, when the entire frame levels are joined at once.

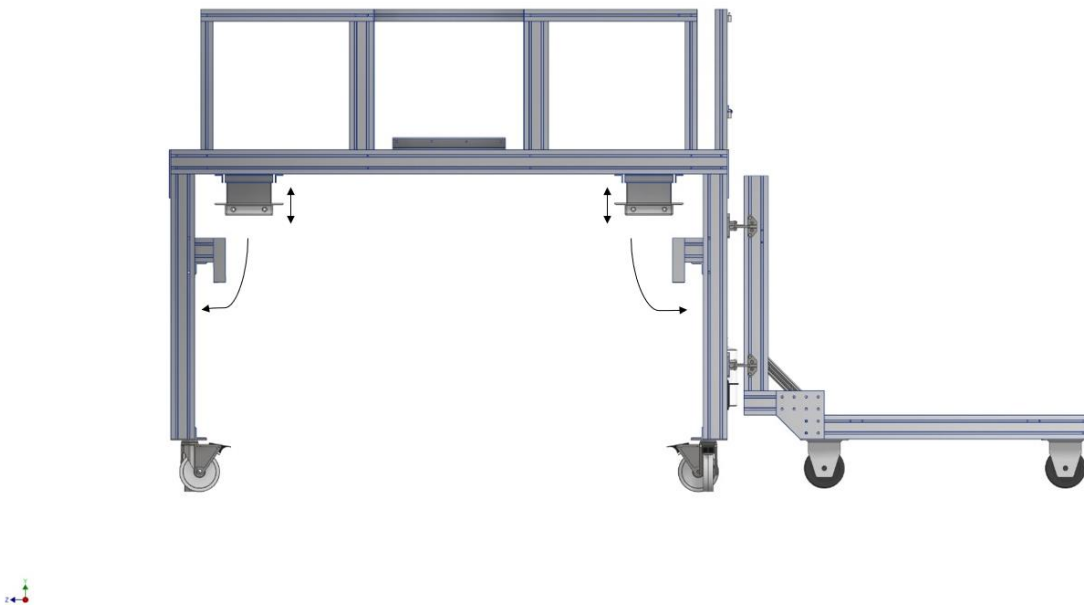


Figure 5.8 RECON framing design side view with articulated features.

Most vertical members are 1515-Lite. Central members secure evaporator and sealing panels. Drop down ducting and case lock in features are visible above and on each side of the computing case position. Range of motion for articulated features is shown through arrows.

The features on the legs of the frame which protrude towards the computing case position are rotating bars with angle profile arms that can be engaged against the corners of the computing case. These features keep the computing case in position when it is mated to the drop

down ducting above, so that the entire system can be moved without damaging ducting structures.

Additional vertical members at the rear support a hinged panel, mounting the electrical control box and 120V receptacles. The hinged feature provides access to the AC control panel behind. Additional framing features provide cable mounting points.

The condenser is mounted across 1530-Lite profiles, similar hinged connections to the main framing as the previous iteration. The base geometry is expanded around the profile of the condenser, increasing the detachable platforms stability. Hinge connection framing is expanded to meet the legs of the main body and provide additional framing which can be used to hold the flexible refrigerant lines.

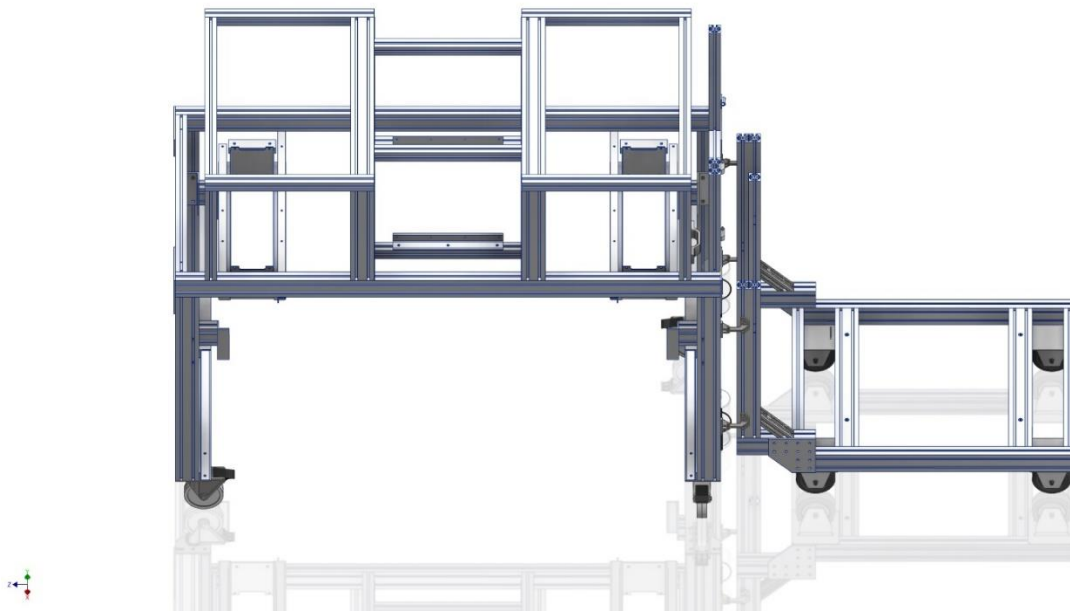


Figure 5.9 RECON framing design render side isometric view.

Framing design view with duct framing features visible. Additional angle extrusions members are used to stiffen duct panels where articulated ducting is mounted.

5.3.2 Cooling and Ducting Design

RECON iteration mounts the same IQ27000V AC system. This iteration places the evaporator within symmetrical framing with expanded ducting chambers. Duct panel material is made of 5052-H32 .09" thickness Aluminum sheet metal. Ducting panel design is simplified, all duct panels are mounted to the outside of the framing structure. Ducting and framing structure are integrated so that the majority of closed-circuit cooling is achieved without welding or folding fabrication required. The rear panel, behind the electrical control box is shaped so that access to the AC control panel and refrigerant lines are open without pass through cutouts required.

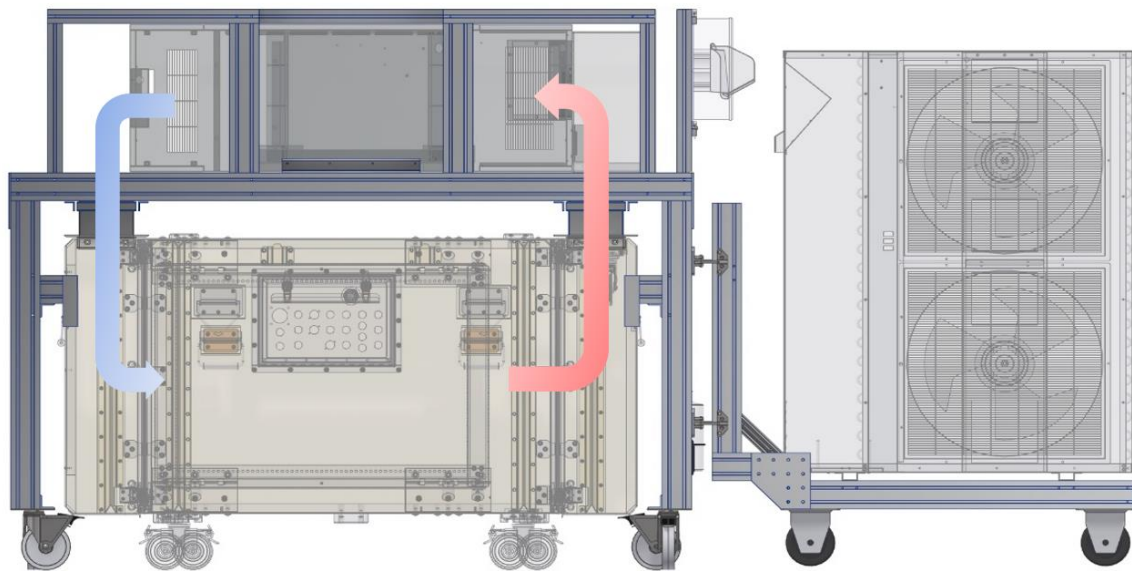


Figure 5.10 RECON cooling design render with airflow direction.

Flow direction through ducting cells. Airflow passes through drop down ducting which mates to the lid cutouts of the Impact case.

Panel sections are sealed using cork $\frac{1}{2}$ " width $\frac{1}{8}$ " thick adhesive strips around the borders of framing-panel contours. Internal framing profiles are sealed using foam inserts placed within the t-slot profiles. Panels mounted outside of the frame allow spacing above and on the sides of the evaporator for 1.5" thick stiff foam board insulation, providing R-7.5 thermal resistivity values for most sections. Insulation is reduced for sections that would obstruct airflow, such as lower panels around the articulated ducting. Overall insulation is set to exceed resistivity values calculated for in cooling requirements.



Figure 5.11 RECON ducting panel design render side isometric view.

Side view of ducting panels, and overhead panels. All evaporator faces are insulated, aside from the articulated ducting and control panel access.

Articulated ducting features engage closed circuit cooling when the computing case is rolled into position below the evaporator framing. Rectangular welded duct section forms a male connection to the case lid cutouts. Sliding surfaces are sealed and held by tubular insulation

strips, applied by angle extrusions surrounding each panel interaction interface. Ducting is manually extended or retracted.

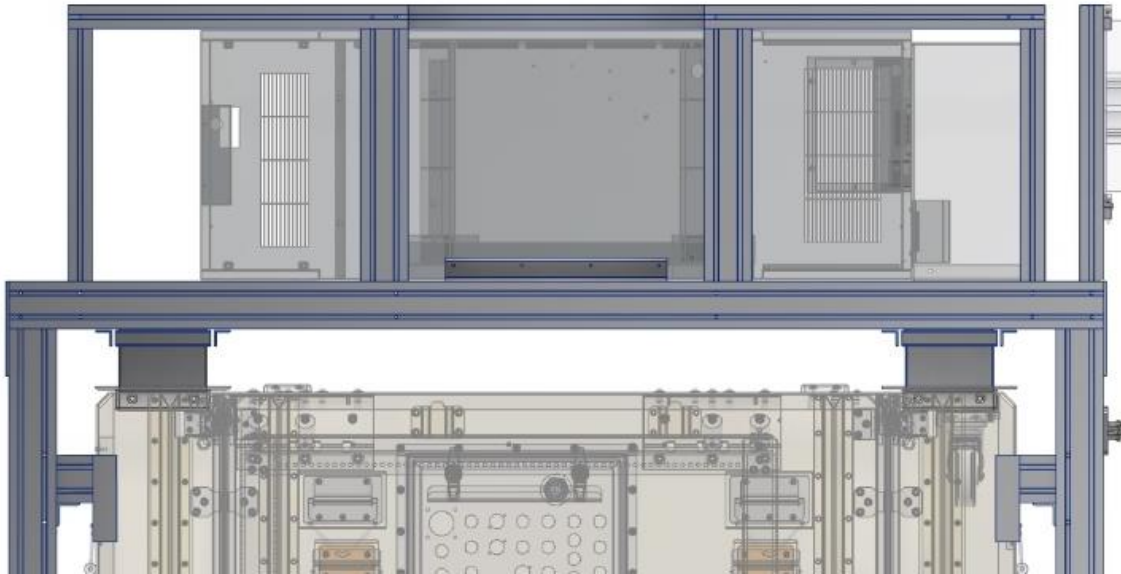


Figure 5.12 RECON design render of cooling component placement.

Transparent side view. Drop down ducting is composed of a rectangular duct section, which is lipped by an angle bracket circumference, providing a sealing surface against the case lid when mated.

5.3.3 Electrical and Control Systems

Electrical control features are mounted at the rear of the main body on a hinged panel. The breaker box splits the 240V power into 120V power for peripheral equipment and 240V power for the computing and AC units. Breakers allow slow burn limits of 60A over two 30A breakers for AC components, and 40A over two 20A breakers for condenser and evaporator units. Two additional 20A breakers limit the 120V receptacles. Cabling remains the same, the control box hosts the power connections supplying the computing components.

AC control interface remains native to the evaporator, accessed at the rear behind the electrical control panel. AC control interface allows control setpoints for cooling target temperature, threshold temperature before condenser activation and deactivation. Settings are adjusted according to performance dynamics of the entire system under the expected heat load of the computing operations.

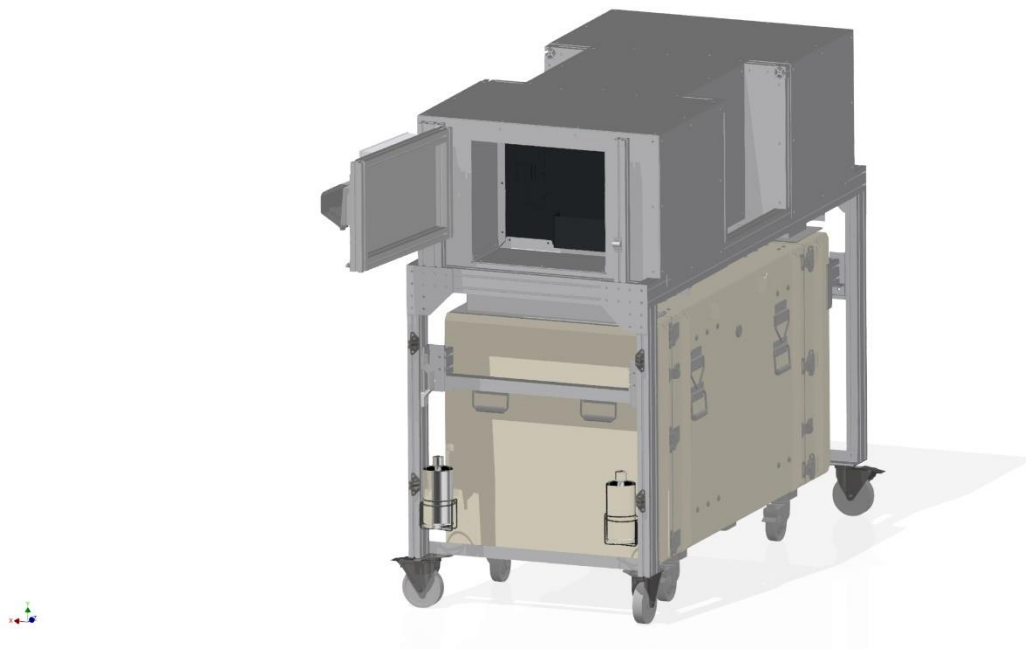


Figure 5.13 RECON design render view of rear control panel access.

Rear duct panel with AC control and refrigerant line connection access points. Condensate containers mounted on rear leg framing. Electrical control panel is hinged in order to provide access at the same location.

5.4 RECON Build

5.4.1 Construction

Two RECON framing/cooling systems were built, and one additional RECON computing stack was built for this iteration. Simplified framing and ducting translated into a simplified

construction process, with reduced complexity and part numbers. Positions of the electrical control box and lower cross members on leg framing are changed in order to achieve better maneuverability in facilities.



Figure 5.14 RECON framing during construction.

Open framing of RECON during build. Fasteners add to the overall dimensions of the assembly, expanding overall width dimensions by $\frac{1}{2}$ " from framing dimensions.



Figure 5.15 RECON duct panel insulation and sealing construction.

Rigid board insulation, foam, sealing tape, and silicon paste are implemented to seal and insulate closed circuit cooling.

5.4.2 As-Built Results

5.4.2.1 RECON

Completed RECON unit with cooling systems are pictured. For testing within facility, the lock-in framing features were removed, however, are installed for transportation and mobile operation. Overall dimensions are slightly larger due to fastener head height. Side joining brackets connecting the legs and main chassis are moved to face forward and backward, elongating the dimensions instead of widening them, maintaining maneuverability through door widths.



Figure 5.16 RECON as-built isometric view.

Lock in framing features removed in picture.



Figure 5.17 RECON as-built side view.

Refrigerant lines and electrical cabling fit between main units.



Figure 5.18 RECON as-built front view.

5.4.2.2 Computing Stack

The RECON computing stack was updated through the development of the cooling and framing systems, adding a supplementary networking switch and uninterruptible power supply. The UPS added electrical protection and regulation against noisy incoming power sources, converting the incoming 240 AC current into DC, then converting that into clean 208V AC power into the PDU and computing stack. The network switch interacts with local networks and fulfills the functionality not covered by the Mellanox switch, managing data flow between the DGX-1 and Dell Precision, as well as pulling data from the DDN storage unit. These components add the intractability and hardening needed to independently operate from a main computing facility.

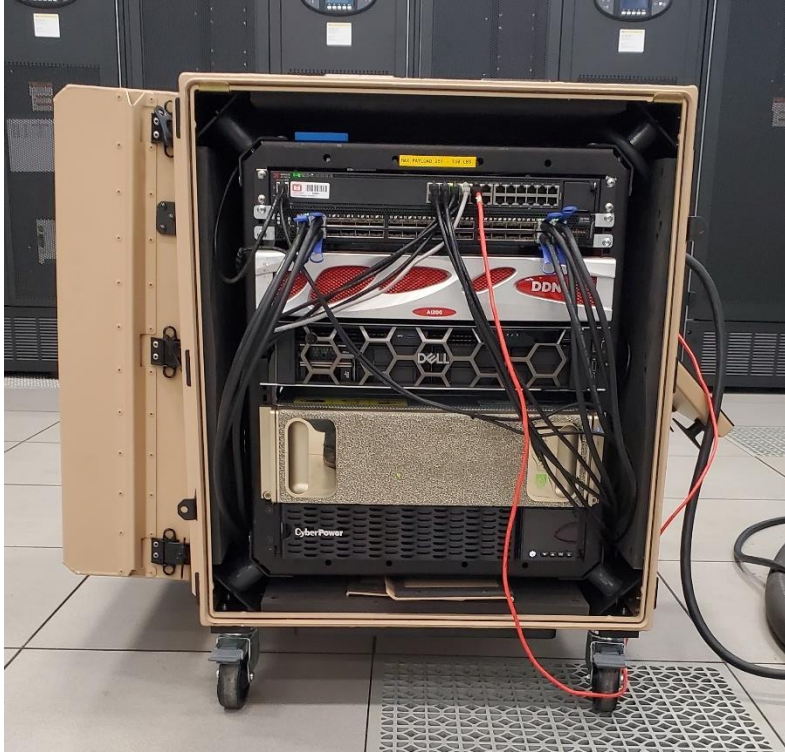


Figure 5.19 RECON updated computing stack.

Updated computing stack adds necessary electrical hardening and data intractability missing from main the HPC pipeline, necessary for an edge node.

5.5 RECON Evaluation and Development

5.5.1 Overview

RECON system evaluation tested component systems and the entire system under different use cases and configurations. The cooling system, computing pipeline, and electrical system are evaluated in different combinations under different use-case environments, revealing the capabilities of the systems individually and cohesively. Design changes and modifications are made as needed in order to maintain functionality.

5.5.2 Evaluation

5.5.2.1 Cooling System Open Circuit

Initial testing of the system started checking both the functions of the AC system and the propane fueled generator. These tests did not run the HPC stack with power connected in order to safeguard from untested electrical faults and temperature variations. In addition to testing the basic functionality of both the generator's capacity to run the AC system, and the AC system's capacity to cool the open or closed-circuit air, the response time and controls native to both systems were tested to see if they could work functionally together. Both AC systems on the two RECON units were tested and worked out for basic functionality, leaks and electrical soundness beforehand. Ambient temperature was 85-90°F in the shade through the duration of the tests, with 40% humidity.

The first test setup shown below ran the AC and generator system at a constant state, taking in ambient air 85°F and returning cool air at 65°F at steady state. The control system within the AC was set to respond at 77°F, given a threshold of 4 degrees, effectively 81°F. Kick on of the compressor, which is one of the peak current draws, was managed by the generator without issue. An open-air setup like this indicated that with the current air flow rate, assuming the processing systems exhaust is roughly the same temperature as ambient, roughly 90°F, the AC system could cool the computing system.



Figure 5.20 Open circuit cooling system test setup.

Cooling system and generator in open circuit, ambient temperature heat load, test conditions, testing steady state ambient air-cooling capability.

5.5.2.2 Cooling System Closed Circuit

The second part of this test connected an empty HPC container to the AC system, creating a closed-circuit airflow, testing the performance of the cooling system under rapid cycling. There was no heat load within the system aside from external conduction, without active computing components. This tested the system's response to rapid temperature variations within the closed circuit, where the AC system, at cooling threshold of 73°F, cooled the circuit until circulatory air through the computing case, computing components and evaporator reached 65°F. The relatively rapid cycling of the compressor, with compressor down time less than 7 minutes, resulted in current draw spikes of 65-70 Amps, disabling the generator. It was found that rapid

cycling of the compressor did not allow pressures across it to equalize in time, causing the current draw. The closed-circuit loop responds quickly to cooling inputs, as the cooling capacity is meant to counteract an active heat output from the computing components. Additionally, the refrigerant system operated outside of the normal range typical pressures and temperatures, where compressor output gave lower pressure and temperature at 94PSI at 28°F and higher pressure on evaporator side at 320PSI at 103°F. This indicated that there was higher flow resistance through the refrigerant system, either due to the flex line diameter and length (1/2" and 3/8" diameter), lack of standard insulation to keep flexibility, or manufacturer error in valve orientation. Additionally, due to the below freezing refrigerant running through the evaporator, evaporator could run the risk of freezing up under long cycle times.

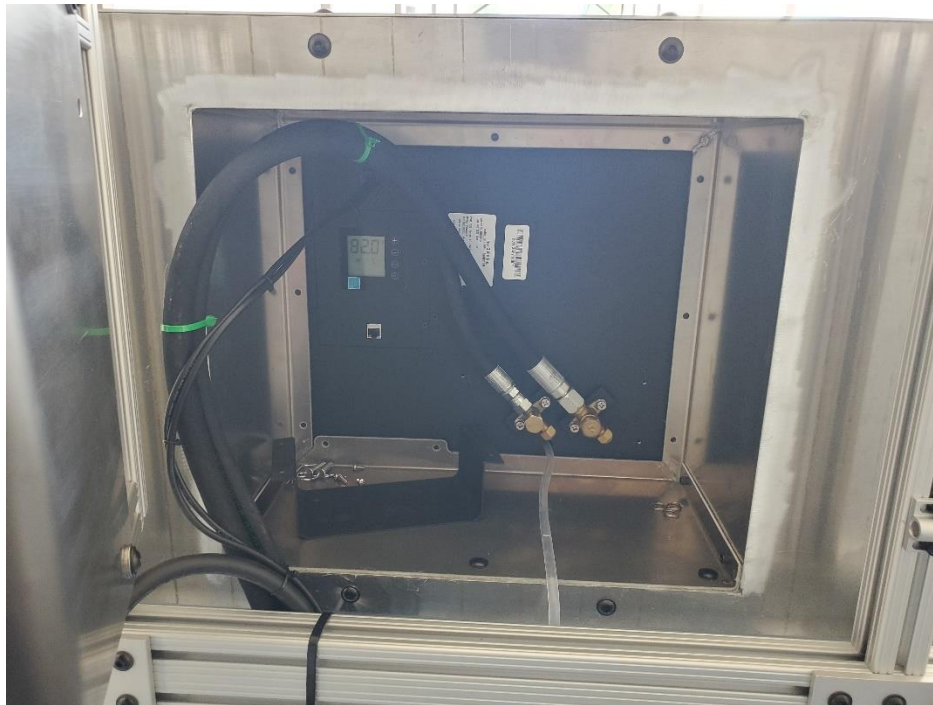


Figure 5.21 Evaporator control unit and refrigerant lines during evaluation.

5.5.2.3 Cooling System Development

Changes to the cooling system, in order to alleviate the previous problems of rapid cycling and excessive current draw, include installation of a time delay switch in line to the compressor, and an increased temperature threshold range before compressor kick on is triggered. The timer addition prevents any signal from the control system from re enabling the condenser until the timer's duration is elapsed. Longer cooling cycle down time also ensures that ice buildup on the evaporator heat exchanger is cleared each cycle, preventing greater failure modes.

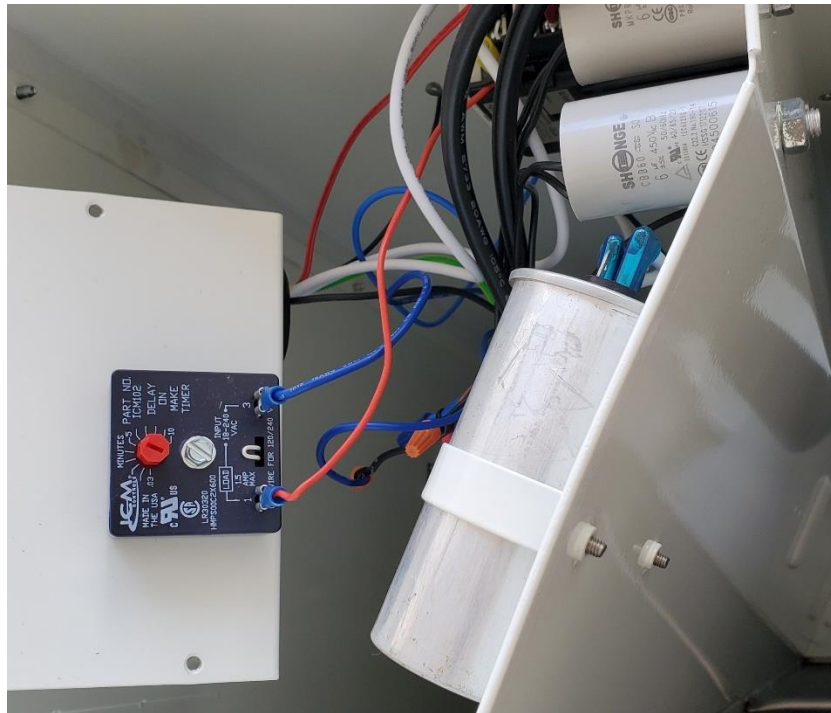


Figure 5.22 Condenser kick-on timer modification.

Delay timer installed inside compressor housing, in line to compressor kick on, sets condition so that compressor will not turn on until set time elapses after compressor shut off. Set to 5 minutes during testing.

5.5.2.4 Computing Benchmark

Computing benchmark testing of the RECON system was done while using established shore power, non-generator power, in order to establish the full computing and cooling capabilities of the RECON system under high ambient temperature conditions, which averaged 85°F. Mellanox switch units have the potential to fail during low power scenarios at start up, prompting changes to the testing and start up procedure. Shore power was selected in order to prevent any electrical system failures under the simulated maximum load.



Figure 5.23 RECON benchmark test for cooling and computing systems.

On-shore power setup for DGX-1 benchmark testing, with full processing thermal load to cooling system. Raritan PDU and benchmark API are accessed through bulkhead connections.



Figure 5.24 RECON computing benchmark test setup.

5.5.2.5 Computing Benchmark Results

This test found that the computing system ran at a power consumption of 1.71KW at idle, and at 3.2KW average during the benchmarking process. The power consumption slowly rose in average by 100W over the course of each 1-hour test due to processing inefficiencies introduced over continues run time. With the extended down time of the compressor between cooling cycles, the system was still able to maintain closed circuit temperatures below a critical threshold of 90°F, internally cycling between 85°F to 65°F over 25-minute average compressor on time to 9 minutes of compressor down time. At an ambient outside temperature of 85°F, in the shade, such a longer compressor up time indicates the AC system is reaching full capacity at lower ambient temperatures than expected, potentially due to the refrigerant line inefficiencies or other factors. This indicates that partial processing load limits need to be applied at ambient temperatures above 90 °F, up to the limit of high-end requirement of 115°F. Overall for this test,

the system API showed that core temperatures were maintained under the critical limits successfully during the test, and the cooling-computing functionalities are successful.

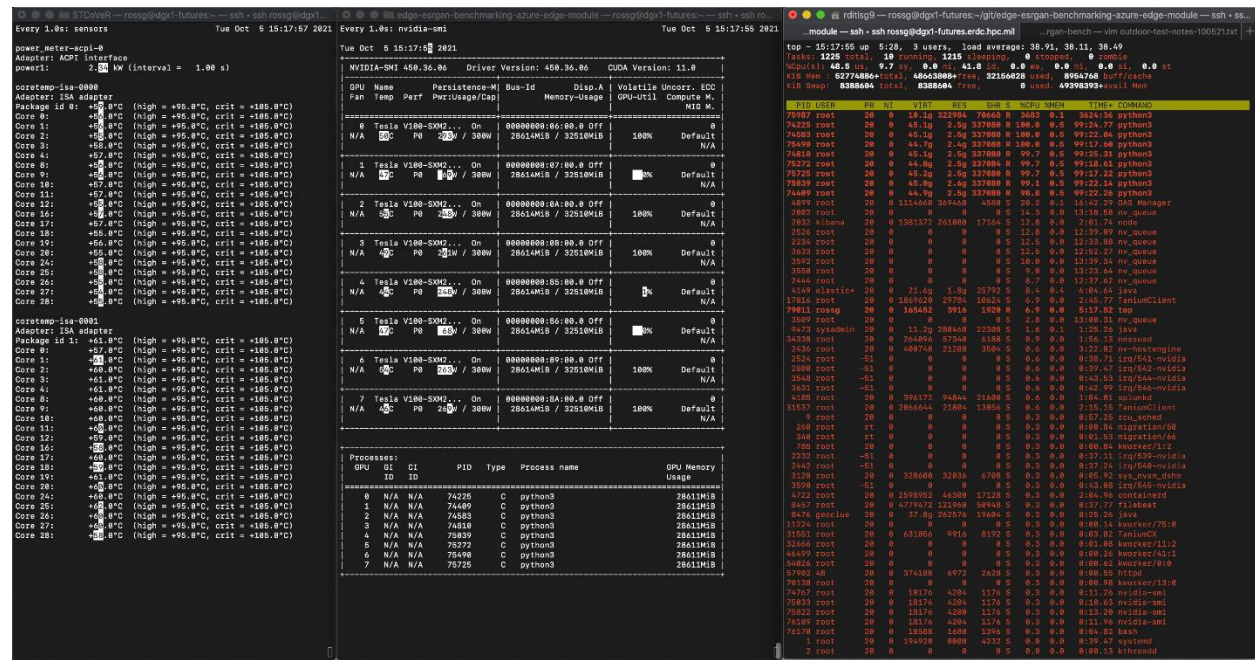


Figure 5.25 Benchmark control API for Nvidia DGX-1.

Benchmark interface showing full running statistics of NVIDIA DGX performance, showing core temperatures and utilization during the benchmark stress test, providing the critical information needed for health monitoring of the system. Cores self-regulate temperature given adequate cool air provided through case.

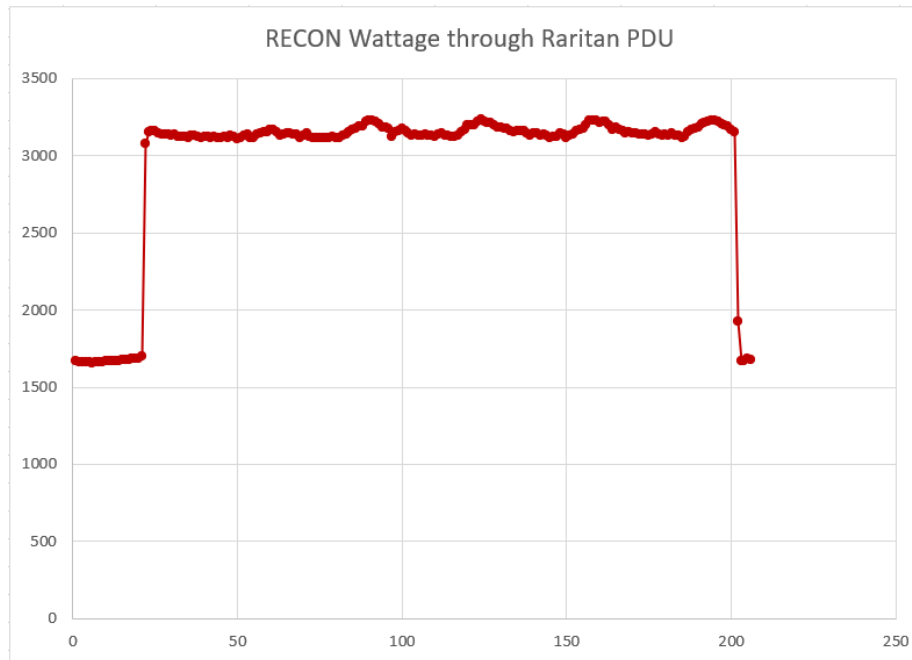


Figure 5.26 RECON benchmark test wattage through PDU.

Raritan smart PDU power log through duration of shore power testing with benchmarking program, in Watts over minute duration. Averages 3.2 kW running while 1.7 kW for idle power consumption, including all computing components running during benchmark test.

5.5.2.6 Electrical and Computing System Development

Testing of the operation of RECON revealed the electrical damage of the system under generator power, additionally revealing that edge nodes require hardening against all avenues of operation when computing components are outside central facilities. The earlier amperage draw spikes from the compressor kick on and cooling system operation, along with startup amperage spikes from component startup resulted in damage to sensitive components. The Mellanox InfiniBand Switch unit of the system experienced component failures on occasions following startup, and operation on generator power multiple times, as far as destroying that specific unit. To correct this issue, the internal computing components were re-arranged to include a rack mounted uninterruptable power supply. This UPS provides the capacitive electrical balance to

the computing system to compensate for the generators momentary inability to provide adequate amperage. This addition allows for full RECON setup testing with generator power to proceed with greater stability toward edge conditions. The UPS is placed at the bottom of the stack, weighing 220 lbs., a significant addition. Changes to the heat flow and computing component weight will change some of the cooling dynamics within the case, as well as the shock and vibration resistance afforded by the 251-350 lbs. tuned payload capacity of the Impact case rack mount dampening system. The components mainly reoriented are minor heat contributors, and the main processing components remain in the optimal position. The new configuration is further evaluated.

5.5.2.7 Mobile Operation

The mobile use case in pursuit of seamless edge HPC applications mounts RECON on a mobile platform, maintaining the required proximity to the edge devices and sensor for uninterrupted service. This configuration tested the operation of the NVIDIA DGX-1 and the supporting components, similar to the benchmark testing conducted on shore power, however also under generator power and on a mobile platform, to run the benchmark program while moving along a specified paved route. The configuration mounted RECON and the 13kW generator on a 20-foot camper trailer, weighing under 2000 pounds with a double axel. The deployed RECON system was strapped in internal to that camper, aside the generator setup, as if operating in a static condition. The condenser fan was oriented to dump heat outside the trailer environment through a side door. All components were secured enough not to displace independently of the trailer movement. Given that mobile operation of the system was previously untested, the route for the test was planned to circle on paved roads without exceeding 10 miles per hour.



Figure 5.27 Mobile operation use-case trailer.



Figure 5.28 Mobile operation use-case RECON setup.

RECON full system setup within the trailer is powered by the generator, running the DGX-1 benchmark test.

5.5.2.8 Mobile Operation Results

The test proceeded on the route for 12 minutes successfully until the conclusion of the test. The temperature was an average 85°F, in varying conditions of shade and full sunlight due to the different cover across the route. There were no apparent issues caused by movement during operation, acknowledging the mild route and speed traveled. The full system took about 1 hour to secure and setup within the trailer, and less than 30 minutes to dismount the trailer during the test. The successful run of this test indicates that the processing components line, the NVIDIA DGX-1, Mellanox IB Switch and DDN hard drive can be tentatively operated in mobile use cases, however, there is no indication that any of the internals of those processors and components are designed for dynamic mechanical load on top of their thermal/processing load, aside from their continued operation. Ultimately, RECON can operate at full processing load within an enclosed trailer, during mobile operation. Combined electrical-cooling-ruggedization system performance of RECON was effective for full load processing in relatively moderate temperature environments.

5.5.2.9 Mobility Use-Case

The RECON computing stack was tested for joint functionality evaluations of an edge sensor reconnaissance application. Using a mobile test configuration, where the RECON computing stack was setup to support the processing requirements of a route reconnaissance sensor array, the new computing stack with a UPS and additional network switch was evaluated. The computing stack within the ruggedized Impact case was secured inside the trailer used in the previous test, behind a towing HMMWV with the attached sensor suite. The sensor data is initially generated and packaged by the sensor array, requiring continuous processing for detection and visualization results. Primary processing and visualization of the data was done by

the RECON computing stack. The objective of this test is to evaluate the effective processing performance of the computing stack with actual sensor data during mobile operation.

5.5.2.9.1 Edge Sensor Application

RECON computing stack was initially configured to process the sensor data prior to the full mobility use case test, using recorded data from the sensor array. Data processing requirements are determined by the variety and quantity of modular sensor inputs. This variety can represent general edge applications for reconnaissance applications, where any variety of advanced sensors require continuous and uninterrupted processing. This test case uses a combination of LiDAR, thermal, and visual cameras, along with global positioning system (GPS) units and an Inertial Measurement Units (IMU) to capture, visualize and detect objects along charted paths. The sensor array initially configured with the RECON computing stack. From the initial test function, power and processing requirements were established.

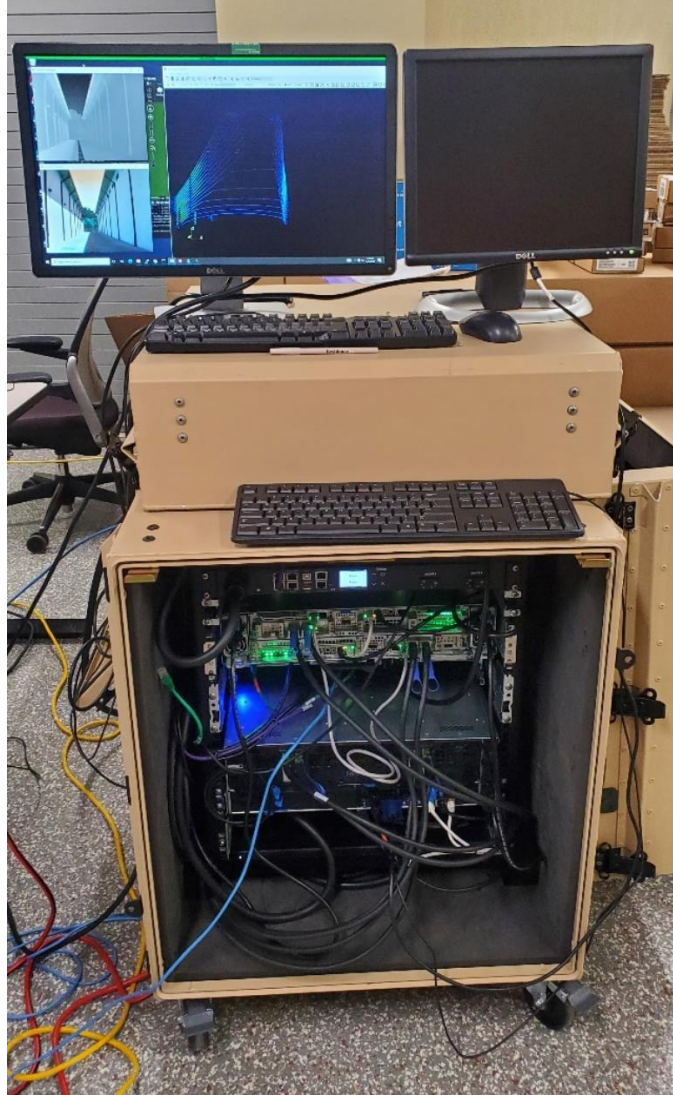


Figure 5.29 Mobility use-case edge sensor initial configuration with RECON.

Initial mobile use case configuration of the RECON computing stack to process and visualize sensor data. Resulting thermal, visual and LiDAR data are visualized in the top left screen pictured.



Figure 5.30 Edge application sensors applied in test configuration.

Sensor array for mobile use case initial configuration, IMU and GPS are on the left and right of rolling table, LiDAR and thermal/visual sensors are in the center.

The sensor array data output is successfully processed by RECON. Sensor data is primarily processed by the DGX-1 GPU, visualized and interacted with through the Dell Precision. The added network switch receives data from the sensor array, and the PDU cleans the input from shore power. PDU power consumption consistently read 2.0kW for the entire computing stack during processing and visualization.

5.5.2.9.2 Mobility Use-Case

RECON computing stack was tested in the full use case test of route reconnaissance. A sensor array was mounted on a lead vehicle, towing the computing stack behind it on a trailer, previously used in the mobility test case. The lead vehicle powered the sensor array and the necessary equipment local to that system, while the RECON computing stack and on-board

cooling system of the trailer were powered by the 13kW generator, mimicking the full RECON system. The generator is mounted on the trailer with propane tanks, supplying 240V to the trailer, and 120V to the on-board AC system. Based on the expected power consumption of 2.0kW, the on-board cooling system can provide the cooling capacity compared to the full RECON cooling system. The trailer provided the mobility and closed environment for the RECON case, so that the computing case would operate with lids removed. This allowed for observation of the rack during mobile operation, as well access to individual components. This test evaluated RECON's computing stability and ruggedization during mobile operation and electrical systems functionality.



Figure 5.31 Mobility use-case test setup.

Route reconnaissance sensor array mounted on lead HMMWV roof with local generator, seen in red. Sensor system passes data for processing to RECON stack in trailer behind, individually powered by the 13kW generator used for RECON testing.

Testing of the full system revealed additional problems with the electrical system. Initial startup saw the shutdown and lockup of the DGX-1, after transitioning the system onto generator power. The generator provided power for the computing system as well as the rooftop trailer cooling system. This cooling system caused frequency changes in the generator power, temporarily shifting power from 60 Hz to 62 Hz, triggering the UPS to temporarily supply power based on its controls. UPS batteries can run the RECON stack for 15 minutes on idle, and 3 minutes on full processing load, however, the frequent shifting of the power further escalated existing faults of a failing battery within the UPS, eventually causing it to degrade and short the system. All components were recoverable after resolution of the UPS. Frequency responses would not have occurred with the standard RECON 240V cooling system in consideration for future test cases.

A system benchmark was then run on the DGX-1 in order to verify full functionality of the UPS with generator power. The Dell Precision on the stack was used to operate the benchmark controls, increasing the number of active GPU cores from 2 to 8, eventually running at peak powers of 3750 Watts, typically averaging at 3400 Watts as indicated through the PDU, converting the generator's 245V at 60-62Hz into 208V, 60.0 Hz power. The UPS was capable of conditioning and supplying power to the computing components.

The route reconnaissance test system conducted three passes over a paved route, with the RECON computing stack continuously processing during the entire duration. The test vehicle drove at most 10 MPH on the route, taking on average 21 minute to complete each pass with a few stops and turns per minute. Outside ambient air temperature started at 90°F during the first pass and rose to 97°F for the remaining passes. Internal temperature of the trailer, and the intake temperature of the DGX-1, started at 77.1°F and ended at 80.8°F at the end of the first run. The

external ambient temperature rose to the average 97°F at the start of the following passes, as did the intake temperature for the DGX-1, starting at 83.1°F, gradually rising to a steady temperature of 87.3°F by the start of the third pass. Power consumption of the stack started at 1950 Watts and rose to an average of 2000 Watts during second and third run. Despite the slight power increase, performance was not affected by the higher intake temperatures, consistent with functionality when enclosed in the full RECON cooling system. The shock and vibration load during the test was already dampened by the trailer suspension, however the system was subjected to 4 Hz repetitive vertical displacement, and broad 1 Hz swaying horizontal motion. Atypical shock due to greater bumps or potholes were visibly dampened when the computing stack was observed displacing out of phase to the original shock relative to the case shell. No shocks or bumps out of the typical trailer response exceeded the rack mount dampening system sway clearance despite the 200 lbs. overcapacity. All components remained functional throughout the final test passes, generating successful route data as required by the sensor array.



Figure 5.32 RECON computing stack mounted in trailer during mobility use-case.

Rack system visibly dampened shock and low frequency displacement experienced through trailer movement.

5.6 Results and Discussion

RECON evaluation determined the systems performance against the requirements of the project. System performance and dynamics are discussed in relation to the objectives set from the onset of this effort.

5.6.1 Cooling System

Cooling system tests focused on the selection and performance of the Ice Qube coolers against the requirement for its ability to cool the HPC system at full load in derated conditions up to 115°F. RECON prototypes were not subjected to that requirement temperature during evaluations. The full prototype successfully cooled the system in 90°F ambient temperatures at highest available exposure conditions. During these tests, the cooling system was able to typically cycle on and off, bringing the full load HPC system down to 62°F from 85°F in 15 minutes of closed loop operation, after an off cycle of 10 minutes brought it back to 85°F again. The ambient temperature was 84°F, 22°F greater than the lowest temperature achieved during operation, where the cooling system had more than enough capability to continue lowering temperatures. With a 22°F differential capability and ample cooling capability remaining, as well as a cycle 3:2 cycle on-off time ratio, the cooling system should be able to handle the 115°F external temperature requirement. Compressor cycle time ratio can be increased, and overall operating temperatures can be increased in order to achieve this requirement, awaiting RECON exposure to those conditions.

Inefficiencies in the cooling system are evident. The HPC benchmark test determining the capability of the system at full processing load also saw sub-optimal pressures and temperatures in the refrigerant system. Assuming manufacturer systems are fully functional, inefficiencies could be introduced by RECON design features such as the flexible refrigerant lines, ducting, HPC component placement and resulting airflow pathing. RECON second iteration refrigerant line were lightly insulated, losing and gaining thermal energy from the environment over the 10 feet of exposed lines, impacting cooling efficiency. Ducting on the second iteration was not designed with ducting complexity which optimized smooth airflow

direction, compared to the first iteration, where airflow was required to make three additional turns out from the air handler before reaching the HPC intakes. Additional flow obstruction occurred in the exhaust and return air side, where rear duct panel controls access restricted airflow to through an immediate 90° turn into a 3” wide cross section length, compared to 6” by 22” lid cutout before it, before passing through air handler filter. Modular design requirements influenced design decisions, where the air handling unit structure and body were left unmodified in order to be easily replaceable, however, removal of air handling unit body panels could have further integrated the cooler into ducting design, increasing air flow volume, increasing evaporator efficiency. Finally, HPC component placement could impact heat exchanging efficiency, where the PDU was placed backward on the rear or the rack. The main processing components control most of the airflow with their internal cooling systems, directing the supply air enough to make the effects of the PDU negligible.

Despite the inefficiencies, the system operates to cool all components after control systems were established to control the cycle time. The cooling system dimensions allowed for the high cooling capacity to be fit through facility doors and vehicles, which were previously unavailable. Ice Qube also offers a higher capacity system within the same packaging, the IQ34000V, which provides 34000 BTU/HR within the same dimensions as the current unit. With or without upgrades, RECON achieves requirements for transportability to edge locations facility maneuverability, modularity, and protection from high temperatures.

5.6.2 HPC System

RECON’s HPC stack was specified to provide a variety of computing resources needed for the processing of data from sensors in edge environments, and provided that capability in all tests conducted. Ruggedization and operation of these components took extra considerations.

The packaging and hardening needed to respond to the facets of operation in edge environments added multiple components and control systems to RECON overall. As discussed in the background of edge computing for small scale deployments, the system needs to provide the functionality that is otherwise provided by the infrastructure of central facilities. The addition of those functionalities fulfilled the requirement for creating a modular system which can provide HPC resources at the edge of logistics, however, weight to the overall system impacts just how resilient, maneuverable and transportable it is in edge environments.

Throughout the tests, the computing stack was resilient to shock and vibration during transportation and operation, both stationary and in mobile use cases. Transportation of the ruggedized case in both DGX only, and CPU and UPS added, computing stacks between and across facilities confirmed a base level of vibration resistance. Both ruggedized computing stacks were transported on the case's wheels over roughly textured epoxy floors at walking pace, passing door transitions and trims protruding less than ½" elevation without ramps at that speed. Similarly, the system was transported over asphalt and concrete sidewalks between facilities. Mobile use case testing drove the system during operation through a 6-mile loop with 10 turns and roughly 10 stops, at an average speed of 10 MPH mounted on a trailer, over paved and gravel roads, transitions between the road styles, and slight grades. The trailer and ruggedized case components additionally withstood road construction related ditches which crossed the road perpendicular to travel, mimicking potholes less than 3 inches in depth during operation. This route was consecutively run 3 times, over 25 minutes each run without ceasing processing operation. Driving and road dynamics induced 4 Hz vertical bouncing and 1 Hz horizontal swaying motions during these runs. The system did not suffer any force related damage, even with the additional of the UPS and network switch, overloading the rated capacity of the Impact

rack mount dampening. The increased weight, 200+ lbs., compressed the dampening system additionally, reducing the amplitude of vibration in the system, however, decreasing dampening of high frequency vibrations. Overall shock amplitude resistance is lowered by the higher weight, as the server rack within the case will more easily displace and impact the shell when sway clearance is exceeded. Despite the required case ruggedization updates required, RECON achieved the shock and vibration resistance needed for transportation to edge environments, and limited mobile operation in edge environments.

Ruggedization and resilience of computing components is less accounted for. HPC component design is not typically accounting for mobile operation or transportation into edge environments. Exposure to electrical noise and temperature variations, amperage limits for generators systems and other unique factors of edge operation are not accounted for by every manufacturer. Components like the NVIDIA DGX-1 are more robust in that regard, providing fail safes on incomplete startups and shutdowns, complex controls for clock speed and cooling of GPU cores, responding to sub optimal conditions presented during RECON's testing. The Mellanox InfiniBand switch, however, was not built with the same resilience, experiencing system failure after startup, where the switch was suspected to require high instantaneous power on startup, otherwise destroying itself functionally. Individual component resiliency still applies and cannot be fully mitigated for edge computing. While that issue was resolved with the additional UPS, the clean environment of a central computing facility needs to be mimicked by the full packaging and ruggedization against electrical, thermal and mechanical interaction. For the edge applications considered for RECON, ruggedization requirements are achieved for HPC operation.

5.6.3 Framing and System Operation

The framing system was able to achieve all functional requirement based on the priorities of design placed on each iteration, and through its functioning in each test phase. The 80/20 framing and hardware solutions allowed for a high factor of safety system without high weight costs and fabrication time. Framing systems are flexible in hardware mounting and modification, providing the interaction points required for system operation, duct and component placement, ruggedization for transportation and operation. The design of the framing system achieved the requirements for creating prototype edge deployable HPC systems, connecting the component systems together to form a maneuverable, transportable and ruggedized computing node.

5.6.4 Mobility Use-Case

The mobility use case tests expanded on the base operations of RECON. While the system was originally intended to be deployed to an edge location for the processing of local sensor data in a static position, the operation of RECON within a mobile environment expands the functionality of RECON to further edge functionalities. The use of small-scale edge deployments provides HPC resources to edge sensor applications. Some applications require seamless and continuous computing resources, and either by operating environment or configuration, require directly connected HPC resources in order to achieve high performance, uninterrupted, and low latency computing operation. The expansion of RECON's use case during the mobility test evaluated the systems ruggedization, modularity and ability to operate on various power sources.

The mobile use-case test also simultaneously evaluated the computing systems electrical capabilities when on shore and generator power. The generator, essentially becoming part of the system, tested the electrical ruggedization of the computing system as a modular system away

from RECON cooling and framing structure, achieving more than the base requirements of modular transportation.

This use case provided the opportunity to evaluate the system against an actual edge application. Use of RECON for route, road and obstacle reconnaissance and detection validated the functionality of the system for general or military engineering support. Use of the computing stack during the test revealed which component were being stressed for the sensor applications and showed how the computing stack can be optimized for sensor enhanced reconnaissance scenarios. Sensor data processing of environments typically relies on GPU core processing, leaning on the DGX-1. Visualization of that data for our control and interaction, as well as operating the computing stack, is managed by the Dell precision. That same function can be handled by much less powerful rack components, which could use less power and weigh less, aiding in overall case ruggedization and transportability. If use cases are tested additionally, the HPC resources provided by RECON could be further optimized.

5.7 Conclusion

The evaluation of RECON was able to determine the system's fulfillment of the requirements established by the RECON's objectives when considered under potential USACE use cases. Sub system design and component selection, arranged under the design of two prototype iterations, was able to fulfill the requirements of providing shock and vibration resistance enabling the transportation of the NVIDIA DGX-1 and supporting components in a ruggedized case to edge environments. Design and selection created a modular system enabling sealed transportation of component systems with appropriate aspects of transportability and protection, enabling the mobility of the system to navigate through 3-foot-wide doorways while deployed, and delivering HPC capabilities to the edge with objective balance attributes of

survivability, mobility and maneuverability. Evaluation of RECON further tested the systems for its ability to process sensor data during mobile operation, validating functionality of the RECON USACE use-case objectives. Further testing is required to evaluate the systems full cooling and operating capabilities while subjected to a maximum of 115°F external ambient temperature. Overall, RECON is capable of delivering HPC resources to edge environments.

CHAPTER VI

CONCLUSIONS

6.1 Objectives Background

Examination of the decentralized computing paradigms, compared to standard central computing, for high performance computing and processing capabilities across the general field of coinciding network and physical domains revealed the motivations and challenges for different scenarios of computing objectives. Specific combinations of objectives and scenarios found overall benefit from edge computing paradigms, including the use of “small-scale edge computing deployments” for general and military engineering sensor applications. Edge computing was considered for potential USACE applications, which established the requirements needed to define the design requirements for a deployable HPC resource capable of operating in edge environments. Edge conditions were anticipated based on potential USACE applications. This thesis explored the design process and evaluations required to develop a Ruggedized Edge Computing Node (RECON), comprised primarily of commercially available components.

6.2 Design Response

Design response to the generated requirements selected the base components necessary for deployable HPC in edge conditions. A GPU based 1000 Teraflop computing stack centered around the NVIDIA DGX-1 and supporting components was configured, refining mechanical and electrical requirements. Commercial Impact rack mount ruggedized cases and Ice Qube

27000 BTU/HR capacity server rack split coolers were selected to fulfill base cooling and ruggedization requirements. 80/20 framing, hardware, and aluminum materials were selected in order to fulfill the remaining requirements, integrating the base components towards the objectives of system mobility, survivability and maneuverability. Concepts integrating the developing component selection were refined, leading to a feasible design concept ready for prototype design.

6.3 RECON Development

Prototype 1 of RECON design background was reviewed in response to the base requirements. Objective priorities regarding attribute traits of mobility, survivability and maneuverability in edge environments influenced design decisions throughout the orientation of base components, framing and cooling system design. Prototype 1 was developed with objectives of ruggedization, compactness and modularity in order to bolster the prototypes ability to be deployed further in edge environments with limited logistics resources, in addition to the base requirements. Evaluation of Prototype 1 successfully confirmed the function of base components cohesively as a deployable ruggedized HPC resource, also revealing which objective priorities needed shifting for better operation and response of RECON in edge environments.

Design narration and evaluation of the second prototype iteration, RECON, redesigned the configuration of ruggedization, computing and cooling components in order to be more maneuverable and versatile in edge environments. Framing design focused on simplifying the previously heavy design, reducing lifting requirements for operation of RECON, adding better modular operation capabilities. The necessary hardening and interactability for edge operation was added to the electrical and computing systems. RECON's evaluation successfully determined its functionality for supporting sensor applications in edge environments, fulfilling

all base requirements, aside from one condition not yet evaluated: maximum processing capability while operating in a 115°F ambient environment. Existing system performance indicates that the objective can be achieved, and additionally, the cooling system otherwise can be upgraded to a higher capacity cooler within an identical footprint.

6.4 Further Research

Additional work can be done to further increase the capabilities of RECON. Design constraints that maintain commercial component structure can be relaxed in order to develop an integrated cooler-case design, greatly compacting framing and ducting into more easily transportable systems. Current RECON framing and ducting introduce constrictions with excessive bends to flow paths for the intake of the air handler, integrated design can open flow paths. Integrated design would better expand RECON functionality into mobile use cases if the dimensions are compact enough to operate within host vehicles or even robotics platforms. Overall increases in system mobility and maneuverability will directly increase the range compatible of sensor application.

Cooling system upgrades can increase RECON's capability to support higher power computing stacks. Existing commercial upgrades to both systems include the IQ34000V higher capacity cooler, and the NVIDIA DGX-2 GPU station. While the IQ34000V could be installed without design modification, the DGX-2 would require a taller rack computing stack, however, Impact cases and the structure can be configured to accommodate the height for 2 petaflops of GPU processing performance, doubling the HPC resource within the same logistics profile, all while still using commercially available main components.

Further design can be applied to address the temperature gain during off cycles of the cooling system, adding capacitive thermal banks to store cooling energy or a cooled medium

while the AC system is on the on cycle, to exchange stored cooling into the closed circuit during the off cycle, which can prevent decreases in performance or internal temperatures from rising above 95°F during off cycles.

Further use of RECON for edge applications will contribute to the better evaluation and standardization of edge computing resources in austere environments, furthering the development of advanced sensor applications for edge engineering operations.

REFERENCES

- [1] Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., Kong, J., & Jue, J. P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. *Journal of Systems Architecture*, 98(December 2018), 289–330. <https://doi.org/10.1016/j.sysarc.2019.02.009>
- [2] Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P., & Nikolopoulos, D. S. (2016). Challenges and Opportunities in Edge Computing. *Proceedings - 2016 IEEE International Conference on Smart Cloud, SmartCloud 2016*, 20–26. <https://doi.org/10.1109/SmartCloud.2016.18>
- [3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A View of Cloud Computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.
- [4] A. Reale, A Guide to Edge IoT Analytics. [Online]. <https://www.ibm.com/blogs/internet-of-things/edge-iot-analytics>, Blog, International Business Machines.
- [5] Satyanarayanan, M. (2017). The Emergence of Edge Computing. *Computer*, January.
- [6] K. Ha et al., “Towards Wearable Cognitive Assistance,” *Proc. 12th Int’l Conf. Mobile Systems, Applications, and Services (MobiSys 14)*, 2014, pp. 68–81.
- [7] A. Li et al., “CloudCmp: Comparing Public Cloud Providers,” *Proc. 10th ACM SIGCOMM Conf. Internet Measurement (IMC 10)*, 2010, pp. 1–14.
- [8] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, “The Case for VM-Based Cloudlets in Mobile Computing,” *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, 2009.
- [9] S. Agarwal, M. Philipose, and P. Bahl, “Vision: The Case for Cellular Small Cells for Cloudlets,” in *Proceedings of the International Workshop on Mobile Cloud Computing & Services*, 2014, pp. 1–5.

- [10] K. Ha et al., (2013) “The Impact of Mobile Multimedia Applications on Data Center Consolidation,” Proc. 2013 IEEE Int’l Conf. Cloud Eng. (IC2E 13), pp. 166–176.
- [11] A. Houmansadr, S. A. Zonouz, and R. Berthier, “A Cloud-based Intrusion Detection and Response System for Mobile Phones,” in Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, 2011, pp. 31–32.
- [12] Bagchi, S., Siddiqui, M. B., Wood, P., & Zhang, H. (2020). Dependability In Edge Computing. Communications of the ACM, 63(1), 58–66.
<https://doi.org/10.1145/3362068>