Abstract—In this paper we address the problem of taking the general notion of sustainability as reported in the Brundtland Report and later in Arrow et al approach and interpreting it from a thermodynamics point of view, with a goal of developing operational principles of sustainability in terms of physical principles that complement economic views.

Index Terms—Sustainability, Thermodynamics, Metrics

I. INTRODUCTION

The term sustainability is used frequently now in many different contexts. For example, in the area of engineering, there have been claims of sustainable products, sustainable manufacturing, sustainable designs and so forth. While sustainability embodies a very serious challenge for human beings, it seems that many of these usages have marginalized the term. It is in the interest of supporting a serious discussion about this concept, that we offer these preliminary thoughts. We are inspired by the work of economists and ecologists embodied in the paper by Arrow et al [1] who, starting with the well known statement of sustainability from the UN Brundtland Report [2] have developed a measurable and workable, (though controversial) criteria for sustainability.

The Brundtland UN Commission statement on sustainability says, “sustainable development is the development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs”. This statement brings up many value laden issues and at first blush seems unworkable. For example, what is a need for one person could be considered excessive consumption for another. Furthermore, who is to speak for future generations and to articulate their needs? In addition, what development means is of crucial importance, in particular does development require growth, and if so, what kind.

Kenneth Arrow and a group of economists and ecologists took on the task of interpreting this statement into an essentially economics framework. Ultimately, they reinterpret this statement in terms of the maintenance of a certain kind of resource base called genuine investment (This to ensure that the rate of change of the aggregate utility based on the total consumption must be increasing or at least being equal to zero.) Genuine investment has three components, 1) Manufactured Capital, including the machines, and factories necessary for production, 2) Human Capital including education, and healthcare, and 3) Natural Capital including minerals, fuels and the services of eco-systems including plants, animals, insects etc. Their resulting criterion is that the aggregate of these three types of capital must be constant or increasing. The strength of this criterion is that it is operational. Economists have devised techniques for calculating these types of capital (although with some difficulties and controversy). In fact, the World Bank issues “The Little Green Data Book” each year, which keeps national accounts for many of the countries around the world in terms of genuine investment or a similar metric. According to this book, the world is still behaving in a sustainable way. In spite of depletions in various Natural Capital resources such as energy and mineral resources and CO2 damage, these are offset by the additions to Manufactured and Human Capital, with a resulting “Adjusted net savings” in 2005 of 7.4 % of GNI [3]. This feature, that Manufactured and Human Capital can be substituted for Natural Capital is however, extremely controversial. This is the so called “weak form of sustainability” and it implies, as Herman Daly has put it “that boats are assumed to be substitutes for fish” [4].

II. RESOURCE MAINTENANCE

We agree with the shift in emphasis in the Arrow paper from seeing genuine investment as a substitute for Natural Capital, to maintaining a certain resource base. Of course, identifying and valuing what is a resource does not come value free either. We can make the best use of thermodynamics however, by focusing on bio-physical resources and using aggregate measures of the second law consequences of the conversion of these resources, such as exergy change and entropy production. By using mass, energy and entropy balances, one can make statements about the changes in exergy (or available energy) of components of the resource base as well as the entropy production associated with these transformations.
However, to start this analysis, the first order of business would be to clearly define the system to be analyzed. This includes selecting the boundaries of the system, the components, parameters and the internal forces of the system.

Two of the simplest and commonly used systems for modeling the problem are: 1) a single, materially closed system, and 2) two (or more) open sub-systems that can exchange materials. Conceptually the closed system is, in the first approximation, a shell containing the biosphere of the earth where human beings live and most biological activity occurs. While this system is materially closed (if any eventual material exchange with the rest of the universe and remaining constituents of earth are neglected), it is open to energy transfers, namely, heat (thermal radiation) from the sun and the core, as well as work effects from the gravitational attraction of the moon. An equally common scheme is one that partitions the world into the so called techno-sphere including people and their industrial activities, and the so called eco-sphere containing the “natural system” including ores, eco-system services and natural products that are used to enhance and sustain the life of human beings. The eco-sphere may further be structured as consisting of biosphere, lithosphere, hydrosphere and atmosphere - all open subsystems experiencing interactions among themselves and with the techno-sphere. These systems are materially open and allow for material transfers between these domains. One could easily envision even more complicated arrangements where specific types of techno-domains and eco-domains are identified and these in turn can exchange materials and energy with each other. This leads us to the question, is there a topology of sustainability? And how does the topology affect the solution? Clearly one wants to subdivide the problem to solve it. But could this subdivision preordain the nature of the solution? A related question then is how to partition the problem, and could a solution be made up of the sum of the parts? These are both very fundamental and very important questions, because engineers frequently want to look at smaller parts of the problem such as products or manufacturing processes.

III. A CLOSED SYSTEM EXAMPLE

To take the closed system example a little farther, we start by drawing boundaries around the resource base that is to be maintained. To be concrete about this, we might start out with the system boundary around the world which contains all life on the surface of the earth extending up into the reaches of the atmosphere that can be affected by human activity (perhaps on the order of 50-100 km) and to the depth of the earth and oceans that are accessible by humans (perhaps on the order of 10-20 km). For our purposes this is a materially closed system. Here we are obviously ignoring volcanism, and interplanetary dust and travel. These are issues that do not concern us here. There are a few advantages to this perspective, and also disadvantages that will be discussed shortly. For example, Ayres uses the consequences of an analysis exercised on a closed system to refute Georgescu-Roegen’s so called “fourth law” (on material resources unavailability on earth) [5]. That is, in general, the materials used by humans on earth are not lost. They can become diluted and/or react with other components of the closed system to lower their exergy, but this problem could be solved by the appropriate application of energy resources.

But there will be a lot going on in this closed system. It is not homogeneous and it is not in equilibrium. It does have certain resources R, which are to be maintained. Note that if the components of this system can spontaneously react with each other, these reactions must involve energy interactions – but these must be across the respective sub-system’s boundary. So, to analyze these interactions some structure must be proclaimed – that is the closed system must be a composite of sub-systems.

The full definition of these subsystems will then allow for the calculation of changes in material integrity, in particular the chemical exergy of the materials components of the biosphere. In the aggregate these spontaneous reactions will result in a net production of entropy and a net change in the cumulative exergy of the chemical components of this system. If this were an isolated system the question would be, at what rate is this resource base degrading. For the closed system however, the long term, so called renewable energy inputs are available to restore the quality of this resource base and even improve on it. This is, of course, exactly what nature does by increasing the chemical and physical exergy of various material systems to support the bio-geo-chemical cycles such as the hydrological cycle, and the carbon cycle. For example the exergy lost by atmospheric condensation of water vapor, rain and flow to the oceans is restored by evaporation and transport powered by the sun. Likewise the chemical exergy lost when carbon is oxidized, be it by autotrophic or heterotrophic respiration or even biomass burning, can be restored when carbon dioxide is upgraded to biologically useful carbohydrates by the process of photosynthesis. From this perspective, it is the use of the renewable resources to restore, and even upgrade, the chemical exergy of certain chemical components of the environment that translates into sustainability.

In particular, we believe that one could look at the rates of destruction vs. the rates of restoration as a principal indicator of sustainability. To be sure this is a vast proposal with many details yet to be explored, but one can get a general sense of where things are going by considering the basic events which affect the chemical exergy of our chemical stores (and any eventual physical exergy involved). These would be: 1) the reactive environment leads to a constant destruction of exergy, this can be accelerated by activity (both human and not) that stirs up materials and brings reactive substances to the surface (e.g. mining, plowing, storms etc.), 2) humans further decrease the exergy of mineral resources by depleting the sources of highest concentrations, but 3) at the same time, humans increase the exergy of minerals and metals for example by refining and smelting, 4) however these human activities are powered almost exclusively by the use of fossil fuels which can result in a net reduction in chemical exergy for the entire system if the carbon cycle becomes unbalanced as it currently appears to be, finally, 5) currently, the primary mechanism for
exergy increase for the closed system is Net Primary Productivity (NPP). We believe that the net effect of these events is an ongoing net destruction of chemical exergy dominated by the huge losses in the carbon cycle.

To see where we are going, and the bio-physical limits on this path, consider a scenario that brings 9 billion people up to the per capita energy used in the United States (350 GJ/yr). This total, about 3.2 ZJ, or 100 TW is almost exactly equal to estimates of total NPP [6,7]. But we already appropriate about 40% of terrestrial NPP, mostly for food [8], and many of us are not at all keen on using the rest of NPP as an energy source. Of course, there are other energy options, but this increase is enormous, and most proposals for how to meet such a demand come with their own set of dilemmas.

Indeed there are claims that it is technically feasible to beat the efficiency of NPP, but these claims are always based upon drawing very narrow boundaries for the accounting scheme. To our knowledge, no one has made a serious proposal that this, in fact, could be the case.

This leads us to perhaps the number one core issue; can technology solve the sustainability problem? In particular, can technology be used to save, maintain or substitute for eco-system services? An equally important question is, can technology substitute for behavioral change on the part of humans?

IV. THE ROLE OF HUMANS

The role of human beings in any sustainability resource accounting scheme can be quite complex. In the weak sustainability criterion proposed by Arrow et al, investment in human education, training and health care are clear candidates to be counted in the resource base. In particular, in both Arrow’s paper and the World Bank’s calculation investments in education are counted in the resource base.

From a bio-physical perspective, a couple of alternatives are possible. At the most elementary level, human beings whether well educated or not, would probably be considered not elements of, but consumers of the resource base. This would be shown quite clearly, if one were to model the resource base as an open thermodynamic system with the humans in the so called techno-sphere and the resources in the eco-sphere. The human contribution in this case would be to improve the efficiency of the conversion of the resources into useful goods and services in the techno-sphere. Additionally, and perhaps more importantly, humans would figure in the development of renewable energy sources to provide these goods and services. At the same time, we must not lose sight of the one over riding driver for this dilemma in the first place-human population growth.

There is an additional perspective that comes from the ecological thermodynamics community that seeks to develop a thermodynamic measure to account for the additional work potential in a biological form that has life [Pastres and Fath, 2009] [8]. This is a question that goes back at least to the booklet produced by Erwin Schrodinger in 1944 entitled “What is Life?” Currently there are researchers who propose an ecological exergy measure which attempts to account for this additional potential in terms of the genetic complexity of the life form. If this perspective were included in the thermodynamic accounting scheme, several issues would need to be addressed immediately including, human population growth, species extinction and net primary productivity. No doubt serious questions about substitutability would also be raised.

V. OPEN SYSTEM EXAMPLE

Energy interactions at the global scale of a biosphere model would lead to great difficulties in specifying causality of events within the complex structure of the closed system even if strong correlations between the set of properties and interactions may be identified. Therefore, a meaningful approach to defining sustainability criteria must be in (1) defining a set of essential interacting open sub-systems that constitute the eco-sphere, (2) identifying the states of these sub-systems in conjunction with established interactions and their properties, and (3) formulating resources flows as material and energy interactions among the sub-systems. In principle, each of these sub-systems may have well defined end states (while at the same time we may not need to be aware of all intermediate states). In general, each sub-system’s state of a global model of sustainability would not be in a steady state. Also, depending on the state and interaction conditions, it is feasible to assume that each sub-system would not be in equilibrium.

Following the above formulated minimum set of considerations, an example of a reasonably simple global model of eco-sphere interacting with techno-sphere that includes humans involving their production base and institutions would consist of four, energy and material interacting, thermodynamically well defined sub-systems, as illustrated at the Fig. 1

This implies that open, non-isolated sub-systems in respective, given states interact through energy and material interactions, thus leading to simultaneous changes of mass, energy, and entropy of sub-systems (the other properties may or may not change as well). The rate of change of mass, energy and entropy can be represented as changes of exergy of each subsystem. This change of exergy may be interpreted as an equivalent change of available energy for sub-system’s interactions expressed in terms of thermodynamically allowable work interactions (but, in the given context, it may not necessarily be a work interaction). The magnitude of this time rate change of non-conserved system exergy may be used as one of the metrics for sustainability of a particular system. This would be justified under the assumption that the available energy expresses an inherent potential of a given sub-system to offer a driving force in interactions needed for the overall system’s change of state. If one assumes that certain aspects of sustainability of each sub-system can be measured by a positive or in a limit equal to zero exergy rate change, it should be clear that such change must be a consequence of both (1) interactions and (2) related sub-systems change of
state (i.e., ultimately would depend on the efficiency of internal transformations). It seems clear that the sustainability of a single subsystem can be neither necessary nor sufficient for establishing the sustainability of a combined set of subsystems or the overall system. However, although exergy is not conserved, it is additive, and certain sustainability features of the interacting subsystems may be measured in terms of aggregate exergy change. Note also that the overall system, in general, interacts through energy (dominant) and material (weak) interactions with the rest of the universe.

This framework allows for determining the exergy rate of supply and the exergy rate of use of both material and energy flows. Hence, one may determine entropy generation (and/or exergy loss) of each of the sub-systems. In case of the technosphere, one may argue, that the loss of energy availability can be correlated through exergy conversion efficiencies to the state of the current state of development through, say, the level of technology development.

For a transient, open system, one can show that the following relationship holds:

\[
T_o \dot{S}_{irr} = - \left( \frac{\partial E_{irr}}{\partial t} - T_o \frac{dS}{dt} \right) - \sum J_i W_i + \sum K_i \dot{Q}_K \left( 1 - \frac{t}{\tau} \right) + \sum_{in} \dot{m} (h - T_o s) - \sum_{out} \dot{m} (h - T_o s) + \sum_{in} \left[ \frac{1}{M} (\mu^* - \mu_o) \dot{m} \right] - \sum_{out} \left[ \frac{1}{M} (\mu^* - \mu_o) \dot{m} \right]
\]

\[
(1)
\]

The left hand side of the equation represents the lost exergy in any of the subsystems of Fig. 1 and/or system as a whole. This loss represents the compounded loss/gain of availability caused by resources removal and/or waste deposition within the subsystems as well as all irreversible transformations within the system. The first term on the right hand side indicates the time rate of change of the availability of the considered open sub system as a consequence of the rate changes of energy and entropy, while the remaining four terms account for exergy interactions caused by both energy (work rate and heat rate) and material flow interactions (in and out). The last two terms represent the rates of chemical exergy carried by the respective mass flow rates.

Let us consider the Hydrosphere and Technosphere interaction as an example of an open sub-system analysis for sustainability. The objective would be to explore how much exergy would need to be invested in the form of energy resources for desalination in order to replenish fresh water withdrawals. The hypothetical scenario involves a worldwide calamity that would prevent use of both surface and ground fresh water. So, under this scenario, the supply must be provided by desalinating the sea water from the oceans but using minimum needed amount of renewable energy resources. To solve this problem one can use the exergy balance equation for an open system that executes separation of pure water from the salts in the saline water and utilizes solar energy with an ideal conversion into the exergy rate. This process takes place between the Hydrosphere and Technosphere and involves an open sub-system that accepts saline and delivers fresh water. Under the scenario discussed, the withdrawal rate would be defined based on the withdrawal rate in the US for year 2000, but projected to 2030 for the whole world. Under this scenario, it would be interesting to determine whether the available sea water as a resource may suffice and what the amount of renewable energy resources would be for the task to be completed. The energy calculation would include three cases: (i) minimum theoretical energy rate needed for separation of the required amount for world supply in 2030 at the rate of US withdrawal in 2000, (ii) the energy rate needed for the same task as for (i) but by using the most efficient, real current technology, and (iii) the energy rate for the same task as in (i) but by using the technology from the end of the past century.

The minimum energy resources use is an available energy (exergy) needed for performing separation of NaCl from H\(_2\)O in saline water. This amount, by definition, is equal to the destruction of exergy in that process, i.e., equal to the product of entropy generation and temperature of the environment. Assuming that no other heat and work interactions exist, in a steady state, entropy balance equation reduces to the equality of \( \dot{S}_{irr} \) and the difference between inlet and outlet entropy rate of the open system for the desalination task. As shown by Gyftopoulos [9], separation of salts from sea water may be considered at any instant of time as a removal of a small amount of mineral constituents of the mixture from a large quantity of sea water, leading to the determination of entropy change in the process as follows

\[
\Delta S = - \left( \frac{M_{H_2O}}{R} \log(y_{H_2O}) \right) = - \left( \frac{M_{NaCl}}{R} \log(1 - y_{NaCl}) \right)
\]

\[
\sim \left( \frac{M_{NaCl}}{R} y_{NaCl} \right)
\]

\[
(2)
\]
where y’s represent mol fractions of the constituents (it is assumed that saline water has 35,000 ppm of NaCl salt in it). The entropy rate change, under described conditions, is equal to the entropy generation rate \( S_{\text{gen}} = \Delta S \). By multiplying the entropy generation rate with the surroundings temperature (assumed to be 25°C) the minimum exergy rate to be supplied to the separation process can be determined. The following data are used. The withdrawal of fresh water [11] in US was 1.3×10^12 kg/day in 2000. If one takes into account that the US population in 2000 was 285M [12] and under the condition of the same withdrawal per capita for the world population in 2030, (9 B) [high variant, 12], the required withdrawal per year would be 1.5×10^16 kg/year. Based on the above discussed open system model, the minimum exergy expenditure (a product of entropy generation and surroundings temperature) would be 1.40966 kJ/kg. So, the exergy rate use for only this task in year 2030 would be ~1 TW. This is, indeed, a substantial amount of energy. More specifically, the projected consumption in year 2030 of each of the energy resource types (oil, gas, coal, nuclear, and other) would be of the same order of magnitude, i.e., between 1 and 8 TW, [13]. One should also keep in mind that the scenario considered here offers the energy consumption for fresh water withdrawal rate based on an ideal, thermodynamically minimum exergy rate for the separation process. This constitutes not the technological limit for a given task at a given instant of time and given level of technological development, but the minimum, theoretically possible amount, regardless of how high the technology level may be. From the sustainability point of view it may be of interest to know that the most advanced desalination processes currently under development require at least 4 times larger exergy use [14]. A standard technology of today would require approximately 8 times more than the most efficient state of the art system [14]. Finally, one may note that the technology of the mid past century would require at least an order of magnitude larger exergy expenditure [15]. The critical step in implementing this approach to resources balancing and availability of resources loss determination is the definition of a state of each sub-system. This step in practical applications of the approach is essential and must involve assumptions controlled by the level of rigor imposed.

Finally, we should stress again that the proposed framework for sustainability of a sub-system and or overall eco-sphere system in thermodynamic terms, i.e., in terms of exergy rate change and/or entropy generation in terms of energy units (the product between the temperature factor and entropy generation) IS NOT a sufficient, but it is a necessary criterion for sustainability. In other words, we are not arguing that one can use a single physical criterion to evaluate sustainability of the ecosphere (that would involve the techno-sphere with complex non-physical interactions). We are just arguing that, equally as one cannot expect that weak sustainability criterion based on the time rate of change of genuine investment in the economics framework to be a single criterion, the physical criterion based on thermodynamics (such as entropy generation and/or exergy destruction) cannot either be considered as a sole criterion. It appears that multiple criteria must be used.

VI. TECHNOLOGY AND ECO-SYSTEM SERVICES

It is commonly believed that technology is an essential part of the solution for achieving sustainability. In this section, we consider this claim from a thermodynamic point of view based on the open system example described in Section V. The purpose of almost all technological and manufacturing activities is to decrease the entropy (increase the exergy) of the technosphere. As per the second law and the exergy balance equation (1), this must result in an even larger increase in entropy (reduction in exergy) of the surroundings. Thus, all technological activities, regardless of whether they rely on renewable or nonrenewable resources cause the entropy of the surroundings to increase. In most technological systems, this entropy increase commonly manifests itself as environmental impact. The current approach of technological development is such that this environmental impact is usually large and negative. Thus, it may be argued that any technological development will result in negative environmental impact, and therefore achieving sustainability via technological development may be impossible [16]. Fortunately, as argued in the rest of this section, thermodynamic insight into ecological systems indicates that shifting the paradigm to consider networks of technological-ecological systems may provide a path to sustainability.

The laws of thermodynamics also apply to ecological systems. Since these systems have sustained themselves for millennia, it is worth considering how they deal with the implications of the second law. Like technological systems, ecological systems also aim to decrease their entropy [17]. Thus, they also cause the entropy of their surroundings to increase and cause environmental impact. However, in self-sustaining systems, this impact is not harmful since intermediate systems have evolved to utilize the exergy loss from other systems to result in an exergy cascade. Thus, although individual activities in an ecosystem are not necessarily efficient – consider the waste created in the Fall by the leaves of deciduous trees, or the highly “wasteful” eating habits of squirrels or monkeys – the overall system is extremely efficient. Individual ecological activities can get away with being wasteful because that waste is put to good use by some other species or activity. In addition, this exergy cascade is such that various activities reinforce each other to result in an autocatalytic system. And all of this runs on renewable energy. Of course, there are situations when this quasi-equilibrium of nature is disturbed by natural or anthropogenic events, in which case, the system has some ability to recover its structure and function or move to another state.

This brief summary of the thermodynamic aspects of ecosystems indicates some approaches that could be used to guide technology development. The ecosystem analogy has already been a motivation for ideas such as industrial ecology and biomimicry. The former field encourages the development of industrial symbiotic networks where “waste = food”, so that the efficiency of the overall industrial ecosystem is enhanced, and its environmental impact is reduced. However, this approach does not go far enough in the
development of self-sustaining networks since it focuses mainly on networks of technological systems.

Ecological systems play a crucial role in supporting all planetary activities by providing essential goods and services. However, the role of ecosystem services is often ignored, even in approaches for enhancing sustainability including industrial ecology and life cycle assessment. Thus, although an industrial ecosystem is likely to lead to an exergy cascade, the exergy loss from such a system is still likely to be large and the system is still not likely to be self-sustaining. Overcoming this important shortcoming requires connecting industrial ecology with ecosystem ecology [18] by considering networks, not just of industrial systems, but of industrial and ecological systems. Such networks could be developed by greater interaction between industrial ecology and ecological engineering [19], since the purpose of the latter is to engineer ecosystems for providing the goods and services needed for satisfying human and industrial needs. Accounting for ecosystem services in life cycle assessment is also essential.

Thermodynamics indicates that the common notion that technology can lead to sustainability may be true, but only if the definition of technology is broadened to consider the supporting technological and ecological systems, since without these supporting services, the entropy increase of the surroundings is very likely to have a harmful impact. The current approach of focusing on a single or handful of technologies such as solar, nano or biomass, is unlikely to lead to sustainability. Thermodynamics can play an essential role in understanding the function of ecosystems in the life cycle of economic products. It can also assist in the development of technological-ecological networks by identifying processes with highest exergy losses and guiding the design of efficient and self-sustaining technological-ecological networks that operate within ecological constraints.

CONCLUSION

In this paper, we attempt to outline some of the issues involved in making statements about sustainable development based on thermodynamics. While it is too early to definitively declare any meta-sustainability criteria based on thermodynamics, we believe that thermodynamics provides the right tools for an aggregate assessment of our sustainability from a bio-physical perspective. That is, the thermodynamics framework does offer a system based, rigorous approach to defining loss of availability of resources. And, thermodynamics offers theoretical limits to assessments of technological advancements and their contribution to solutions of sustainability problems.

Although, we have not dwelled on the available data, and only offered a few examples here, we believe that the bio-physical evidence is quite contrary to the neo-classical economics conclusion; we quite clearly are not behaving in a sustainable way at present. At the same time, we fully admit that thermodynamic metrics and derivative criteria must be only a complementary set of metrics and/or criteria for an assessment of sustainability. No single metric (regardless how well aggregated) or derivative criterion is able to offer a completely satisfactory solution for all situations.

Finally, while we have only taken the first steps in considering the partitioning of the sustainability problem, it is already quite apparent that claims about “sustainable products” and other small pieces of the problem are vacuous without a credible connection to a convincing world view of sustainability.

REFERENCES


[12] UN Department of Economic and Social Affairs; http://esa.un.org/unpp/p2k0data.asp


[14] Siemens technology as reported in Siemens Research and Development, Ref. No. RN 2008.11.01e


