Abstract—This paper presents a novel energy-aware and temperature-aware programming model with first-class support for sustainability. A program written in the new language, named Eco, may adaptively adjusts its own behaviors to stay on a given (energy or temperature) budget, avoiding both deficit that would lead to battery drain or CPU overheating, and surplus that could have been used to improve the quality of results. Sustainability management in Eco is captured as a form of supply and demand matching, and the language runtime consistently maintains the equilibrium between supply and demand. Among the efforts of energy-adaptive and temperature-adaptive systems, Eco is distinctive in its role in bridging the programmer and the underlying system, and in particular, bringing both programmer knowledge and application-specific traits into energy optimization. Through a number of intuitive programming abstractions, Eco reduces challenging issues in this domain—such as workload characterization and decision making in adaptation—to simple programming tasks, ultimately offering fine-grained, programmable, and declarative sustainability to energy-efficient computing. Eco is an minimal extension to Java, and has been implemented as an open-source compiler. We validate the usefulness of Eco by upgrading real-world Java applications with energy awareness and temperature awareness.

I. INTRODUCTION

Two trajectories appear to shape up the evolution of modern computing platforms: they become increasingly “nimble,” or increasingly “mighty.” “Nimble” computers—such as RFIDs, wearable electronics, smartphones, tablets, and laptops—often operate on battery with a limited budget. “Mighty” computers—represented by data centers, cloud servers, and many-core clusters—not only consume a large amount of energy, but also face grave reliability concerns due to excessive heat dissipation, with cooling taking up to 35% of their operational costs [3]. Effective energy and thermal management is a critical goal demanding solutions from all layers of the compute stack. In recent years, there is a growing interest in addressing these challenges from the perspective of application software (e.g., [16], [5], [21], [35], [25], [17], [33], [34], [29], [32], [20], [39], [4], [36], [10], where energy and thermal efficiency is pursued and achieved throughout the software lifecycle.

In this paper, we focus on an essential design goal in energy and thermal management: sustainability. A sustainable program is able to “stay on budget,” neither yielding deficit nor surplus. In the context where “budget” refers to available battery — e.g., 20% battery of a smartphone — sustainability entails energy awareness: the sustainable program makes the best effort to not exceed 20% of battery usage (no deficit), while at the same time producing as high-quality results as possible (no surplus). In the context where “budget” refers to the allowance of temperature increase — e.g., 10°C — sustainability entails temperature awareness: avoiding overheating while producing the best results. In essence, sustainability captures a recurring theme in energy and thermal management: balancing the trade-off between energy/thermal budget and the quality of computational results.

A. Eco for Sustainable Programming

Our concrete proposal is Eco, a novel and simple programming model to help developers construct sustainable software by design. As we look forward, we believe programmers have a unique role in application-level energy and thermal management, and the future of software on “nimble” and “mighty” computing platforms belongs to applications that neither “squander” nor “skimp” on resources.

The key insight of the Eco design is that sustainability can be achieved through the negotiation between supply and demand, ultimately maintaining the equilibrium of the two in a fashion similar to market economics. Syntactically, Eco is a minimal extension to Java, taken with two important programming abstractions:

- **Demand/Supply Characterization**: an Eco programmer may associate a fine-grained program component of interest (the “demand”) with a budget (the “supply”), through a new construct called **sustainable blocks**
- **Adaptability Support**: an Eco programmer may define alternative application behaviors which may achieve the same application-specific goal but with different quality levels, through a principled and fine-grained construct called **first-class mode cases**

Throughout the execution of the code enclosed in the sustainable block, the Eco program runtime continuously monitors the change in demand and supply. As demonstrated in Figure 1, the sustainable execution should have the remaining demand (the solid line) closely matches the remaining supply (the dotted line). If demand far exceeds supply at a snapshot of...
the execution, the program adapts to program behaviors that may produce results in lower quality. If supply far exceeds demand, adaptation in the opposite fashion happens.

In a nutshell, the design scope of Eco subsumes programming abstractions to capture the essence of sustainable programming, compiler inference algorithms to automatically identify program points for monitoring and adaptation, and runtime support to efficiently maintain the balance between demand and supply.

B. Contributions

To the best of our knowledge, Eco is the first programming model to provide first-class support for characterizing, negotiating, and maintaining budget in energy-aware and temperature-aware applications. This is also the first time fundamental concepts in market economics—supply and demand—are crystallized as programming abstractions. In the context of existing energy/thermal budget management systems [12], [38], [40], [9], [11], [13], [27]—most from VLSI, architecture, and OS communities—Eco is distinctive in its attempt to bridge application software with underlying systems, for at least three distinct reasons:

1) it elevates predominately system-level concepts such as energy budgeting and temperature control to the forefront of application software development, through supply characterization.

2) it taps the knowledge and insight of programmers to automate fine-grained energy and thermal management, through demand characterization.

3) it brings green-conscious programmers into the equation of energy and thermal optimization, through principled programmer-defined adaptability support.

Eco is a rigorously defined language [41], but more importantly, it is also a practical and expressive programming model for developing sustainable real-world applications. This paper places emphasis in highlighting Eco’s role in green software engineering: how it empowers programmers with battery and temperature awareness, how it helps balance the trade-off between energy/thermal efficiency and quality of service, and what the recurring programming patterns in sustainable software are. Eco is developed as an open-source compiler:

https://github.com/pl-eco/ECO

In summary, this paper makes the following contributions:

- It introduces a simple programming model to provide fine-grained sustainability through supply/demand characterization, program transformations to automate and optimize sustainability management, and a program runtime design to allow for supply/demand negotiation.
- It provides first-class support to adaptive software. The design considers the consistency in adaptation, and seamlessly integrates adaptability and sustainability through compiler-directed automatic calibration and adaptation.
- It describes a set of common programming idioms, which may serve as a first-step toward understanding programming patterns in sustainable programming and constructing energy-aware and temperature-aware software.
- It demonstrates the feasibility and effectiveness of sustainable programming through extending real-world applications with the ability of battery awareness and temperature awareness, enabled by an open-source compiler.

II. SUSTAINABLE PROGRAMMING BY EXAMPLE

We use a simple ray tracing example in Figure 2 to demonstrate the basic ideas of sustainable programming. Raytracing is a technique in rendering where an object in the form of a Model is rendered into an Image—a 2D array of colored pixels—based on the Rays cast from the Camera to the object. Eco is an extension to Java, with several notable features we highlight next.

- **a) Sustainable blocks**: An energy-conscious programmer may declare the most expensive part of her program in sustainable blocks. In this example, the block encloses the

```java
class Raytracer {
    mcase<Res> res = {
        lo: new Res(640, 480);
        mid: new Res(800, 600);
        hi: new Res(1024, 768);
    };

    mcase<int> depth = {
        lo: 10;
        mid: 20;
        hi: 30;
    };

    Ray[][] getView(Camera camera, Res res) { ... }

    Image[] run(Camera camera, Model[] models) {
        Image[] images;
        int num, IMAX, JMAX, IMGS = 100;
        sustainable {
            for (num = 0; num < IMGS; num++) {
                uniform {
                    Ray[][] view = getView(camera, res);
                    Image image = new Image(res);
                    IMAX = res.x;
                    JMAX = res.y;
                }
                for (int i = 0; i < IMAX; i++) {
                    for (int j = 0; j < JMAX; j++) {
                        image.setColor(i, j, view[i][j].trace(models[num], depth)
                    );
                }
                images[num] = image;
            }
        }
        bsupply (0.2 + battery)
        demand (IMGS) -> (IMGS - num);
        return images;
    }

class Model {...} // model to be rendered
class Camera {...} // camera for rendering
class Image {...} // 2D array of color pixels
class Ray {...} // ray
class Res {...} // resolution

Fig. 2. A Battery-Aware Raytracing Application
```
ray tracing for IMGS number of images, as shown between Line 22 and Line 41. The programmer can set a budget for executing this code fragment through supply characterization: in Line 40, the budget (also known as supply sum) is set at 20% of the remaining battery. To help the language runtime correlate the change in battery over time and the progress of the program execution, Eco programmers are asked to provide a “progress indicator” through demand characterization. In this simplified example, the programmer just says that the demand sum—i.e., the overall number of units of work—is IMGS (the number of images), and at a given moment during the sustainable block execution, the demand gauge—the indicator for the remaining number of units of work—is dependent on the loop index num, i.e., through expression IMGS - num. Based on supply characterization and demand characterization, Eco introspects

- the proportion of remaining demand, i.e., demand gauge relative to demand sum and
- the proportion of remaining supply, i.e., remaining battery supply relative to overall battery supply

at various automatically determined “checkpoints” during the execution of the sustainable block, and deem the execution: (i) sustainable if the two are on similar levels; (ii) deficit if the former significantly exceeds the latter; (iii) surplus if the latter significantly exceeds the former.

b) Modes and First-Class Mode Cases: Whenever deficit or surplus happens, the program adjusts itself, guided by specifications of modes and first-class mode cases.

Borrowed from a previous work [10], modes are a discrete characterization of energy adaptation states. In Line 1 and Line 2, the program defines three modes, lo, mid, and hi, indicating the program may function under three different states of adaption. In Eco, modes form a total order, with lo/mid/hi in the example aligned with the intuition that the resulting code is likely to consume the least/average/most energy and produce results of the lowest/average/highest quality, respectively.

Modes matter in Eco because they play a central role in the design of mode cases. In a nutshell, a mode case is a data-centric way to define alternative behaviors in the presence of energy adaption. In the example, two mode cases are defined as fields: (i) the res field in Line 5 that says the Raytracer may use the resolution of 640×480 when program is executing the lo mode, 800 × 600 in the mid mode, and 1024 × 768 in the hi mode; (ii) the depth field in Line 11 that says the depth of Raytracer—a key parameter to control the quality of raytracing—may be either 10, 20, or 30, depending on the mode of the executing program. Mode cases in Eco are first-class values: they can be assigned to variables, passed around as method parameters, or stored in fields.

To tie everything together, the executing sustainable block “downgrades” the mode of its execution whenever deficit happens—picking a precedent in the total order—and the mode case used in that block can be intuitively viewed as automatically reduced to the value associated with the new mode, a step we call mode case selection. For example, the mode case of depth under the execution of mid mode would be selected into 20. The scenario for surplus is similar, except that the sustainable block “upgrades” its mode. The described behavior here is aligned with our intuition of sustainability: the program adapts to the less energy-consuming behavior when it runs into a deficit and to the more energy-consuming behavior when it runs into a surplus.

c) Uniform Blocks: Coming with the flexibility of first-class mode cases is a design challenge: allowing for mode case selection at arbitrary time may yield results unacceptable to programmers. For instance, had we allowed the image resolution to change half way through raytracing an image, the program may produce a picture such as Figure 3, whose uneven rendering may unsettle (some) programmers.

An Eco programmer may choose to declare a block of code as uniform, as in Line 24 to Line 29. This precludes mode change from happening during the execution of said block. This consistency-preserving construct is reminiscent of the property of atomicity [15], [22], [19].

d) Battery-Awareness and Temperature Awareness: The example here is an instance of battery-aware programming. With minimal changes, the same program can turn into a temperature-aware one: by simply changing Line 40 to tsupply(10), the same sustainable block is to be executed with a (temperature) budget of 10°C, i.e. not raising CPU temperatures by more than 10°C. The similarity and symmetry of battery-aware programming and temperature-aware programming in Eco do not come as a surprise. They are unified under the same philosophy: just like the allowance of battery usage during a program execution can be viewed as a “budget,” the allowance of temperature raise during a program execution can also be viewed as a “budget.” As a consequence, demand/supply-style characterization and programming are applicable to both goals, and the unified programmer view is an attractive trait to improve programmer productivity.

Indeed, the unification of battery-aware programming and temperature-aware programming goes deeper than syntactical similarities. As we shall see in the next section, the bsupply and tsupply constructs are both syntactical sugars of a more general programming abstraction, and the two hence share the same semantic definition.
III. THE ECO DESIGN: SYNTAX AND RUNTIME SUPPORT

A. Syntax Overview

Just like Java, an ECO program consists of a sequence of classes, with one containing the entry method main. In addition, an ECO program allows for top-level mode declarations in the form of mode \texttt{m} \leftarrow \texttt{m}. We require the \texttt{\leftarrow} relation defined through these declarations forms a total order. We further call the least element and greatest element in the total order the \textit{least mode} and \textit{greatest mode} respectively. Implicitly parameterized by these top-level mode declarations, unary operators \(\uparrow\) and \(\downarrow\) are standard successive and precedent operators over total order. Furthermore, we define \(\uparrow m\) as \(m\) if \(m\) is the greatest mode and \(\downarrow m\) as \(m\) if \(m\) is the least mode.

The core expressions of ECO are defined as follows:

\[
e := \begin{array}{ll}
\text{uniform}\{e\} & \text{expression} \\
| \{e_1; \ldots; e_n\} & \text{expression} \\
| \text{sustainable}\{e\}(e) & \text{characterizer} \\
| \text{select}\ e & \text{characterizer} \\
| \text{calib}\ e & \text{Java expressions} \\
\end{array}
\]

For simplicity, we treat expressions as statements. Expression \(\text{uniform}\{e\}\) defines a uniform block whose body is \(e\). We explain the rest of the syntax shortly. A type in ECO may take three forms: 1) a primitive type (e.g., integer, float), which is abstractly represented as \texttt{num} for simplicity here; 2) an object type \(X\), where \(X\) is the name of a class; 3) a type for the mode case value \texttt{mcased}\(\tau\), which says that the mode case has case members whose types are \(\tau\).

B. Runtime Behavior Overview

The unique runtime features of ECO center around the execution of a sustainable block. There, the (sum and remaining) supply and the (sum and remaining) demand are monitored through checkpoints, each of which we call a step of \textit{supply/demand calibration}, or \textit{calibration} for short. This leads to four interconnected questions at the center of ECO design:

- **Q1**: How to calibrate?
- **Q2**: How should the program adapt based on calibration (through mode case selection)?
- **Q3**: When to calibrate?
- **Q4**: When to select mode cases?

We answer Q1 in Section III-C3 and Q2 in Section III-D. To decouple Q3 from Q1, and Q4 from Q2 for presentation purposes, we introduce two convenience expressions to make calibration point and mode case selection point explicit, \texttt{calib} \(e\), and \texttt{select}\( e\). The first expression says that calibration should be triggered after the evaluation of \(e\), and the second expression says that the mode case represented by \(e\) should be selected. In other words, when we answer Q1 and Q2, we circumvent the “when” questions by assuming they are explicitly “marked” in the program. In Section IV, we will introduce several program transformations to answer Q3 and Q4, i.e., automatically inserting \texttt{calib}\( e\) and \texttt{select}\( e\) expressions to the ECO source program.

C. Syntactical and Runtime Support for Sustainability

1) Syntax and Encodings: The general form of a sustainable block in ECO is as follows:

\[
\begin{array}{c}
\text{sustainable}\{\ldots\}\{e_1; e_2; e_3; e_4\} \\
\quad \ldots \text{// code in the sustainable block} \\
\quad \text{supply}\ (e_1) \to (e_2) \\
\quad \text{demand}\ (e_3) \to (e_4)
\end{array}
\]

where \(e_1, e_3, e_4\) are the expressions for representing the \textit{supply sum}, \textit{demand sum}, \textit{demand gauge} we informally introduced in Sec. II, respectively. The additional expression \(e_2\) is called the \textit{supply gauge}, an expression to indicate the remaining supply. In this presentation, we also abbreviate the sustainable block above as \texttt{sustainable}\{\ldots\}\{\langle e_1; e_2; e_3; e_4\rangle\}, and further call \(\langle e_1; e_2; e_3; e_4\rangle\) the \textit{characterizer} of the sustainable block.

Indeed, the sustainable blocks with \texttt{bsupply} and \texttt{supply} keywords introduced earlier are syntactic sugars, which can be encoded as in Fig. 4. As we shall see in Sec. V, ECO programmers can account for flexible forms of supply variability through supply gauge programming.

2) Sustainable Stack: A ECO program runtime maintains a key data structure called a \textit{sustainability stack} (represented by \(S\)), or \(S\)-stack for short. The \(S\)-stack maintains the context of sustainable/uniform blocks: at a runtime state, each entry of this stack intuitively says that the current encountered program point is \textit{dynamically scoped} in the sustainable/uniform block represented by the entry, with the top entry representing the
most recently encountered scope. There are two important features to highlight from this design:

- Eco supports sustainability management across lexical scopes.
- Eco supports nested sustainable blocks, and nesting between sustainable blocks and uniform blocks.

The first support is a necessity for object-oriented languages: object references can escape lexical scope, and as a result, the program point of calibration may not be lexically enclosed in the sustainable/uniform blocks. In the technical report [41], we provide a detailed example to illustrate this technicality. The S-stack design flexibly enables sustainability management for all program points reachable from a sustainable/uniform block.

The support of nested blocks is important for two reasons. First, this facilitates modular programming, e.g., in a scenario where the designer of a library API declares a sustainable block within the API method, and at the same time the client program using the API may declares its own (nesting) sustainable block. More importantly, nesting support also enables flexible sustainable programming where battery awareness and temperature awareness is required at the same time.

Structurally, each S-stack entry can take one of the two forms (1) a two-element tuple in the form of \((e : m)\), where \(e\) is the characterizer expression of the sustainable block, and \(m\) is the mode of the current sustainable block execution, called the operational mode. (2) a constant element \(\bot\), indicating the entry representing a uniform block.

When the program bootstraps, we place an initial entry to the S-stack, whose operational mode is the greatest mode in the mode specification (a total order). This default treatment is aligned with our intuition that code not subject to sustainability management is the characterizer expression of the sustainable block, and \(m\) is the least mode among \(i\) is the characterizer, and \(m\) is initially the same as the “current” operational mode before the block is entered, where the notion of “current” is defined by the following function:

\[
\text{MNOW}(S) \overset{\text{def}}{=} m \quad \text{if} \ S = S', (e : m)
\]

\[
\text{MNOW}(S) \overset{\text{def}}{=} \text{MNOW}(S') \quad \text{if} \ S = S', \bot
\]

This design is useful in the presence of nested sustainable blocks, where the inner block starts operating at a mode sustainable for the outer block.

3) Sustainability Management as Supply/Demand Matching: Upon the evaluation of the calib expression, there are two possibilities:

- If any entry in the S-stack is \(\bot\), the calibration is a no-op.
- Otherwise, supply/demand matching will be performed for all S-stack entries, with the top item being performed first. The detailed matching process will be explained next. The definition supports nesting, where sustainability management applies to all blocks, with the most immediate (inner) block first.

For each instance of supply/demand matching, the supply sum expression, the supply gauge expression, the demand sum expression, and the demand gauge expression will in turn be evaluated to a numerical value. Let us denote them as \(n_1, n_1', n_2, n_2'\) respectively. The matching between supply and demand is captured by the following predicate, where \(\epsilon\) is a small positive numeric constant close to 0:

\[
\text{MATCH}((n_1; n_1'; n_2; n_2')) \overset{\text{def}}{=} \begin{cases} \frac{n_1'}{n_1} - \frac{n_2}{n_2'} \leq \epsilon \\ \frac{n_1'}{n_1} - \frac{n_2}{n_2'} > \epsilon \end{cases}
\]

\[
\text{SURPLUS}((n_1; n_1'; n_2; n_2')) \overset{\text{def}}{=} \frac{n_1'}{n_1} - \frac{n_2}{n_2'} > \epsilon
\]

\[
\text{DEFICIT}((n_1; n_1'; n_2; n_2')) \overset{\text{def}}{=} \frac{n_1'}{n_1} - \frac{n_2}{n_2'} > \epsilon
\]

Eco language implementation selects \(\epsilon = 0.1\) by default: the program is considered “sustainable” if the remaining proportions of supply and demand are within \(\pm 10\%\) difference.

D. First-Class Mode Cases

Expression \(\{m_1 : e_1; \ldots ; m_n : e_n\}^T\) represents a mode case, where each element \(m_k : e_k\) (for \(1 \leq k \leq n\)) in the sequence is called a case member. At run time, the expression associated with each case member is evaluated at mode case definition time. As a result, each case member of a mode case holds a value at runtime.

First-class mode cases at run time are selected based on the operational mode. For now, to circumvent Q4, let us assume all mode case selection points are explicitly identified, through the select expression. The expression select \(e\) at run time evaluates to \(v_i\) following this definition:

\[
\{m_1 : v_1, \ldots , m_n : v_n\}/m \overset{\text{def}}{=} v_i
\]

where \(e\) evaluates to \(\{m_1 : v_1, \ldots , m_n : v_n\}\), \(m\) is the operational mode, and \(m_i\) is the least mode among \(\{m_1, \ldots , m_n\}\) that is equal to or greater than \(m\) according to the total order defined by the program.

Recall that the program bootstraps with the greatest mode. This implies when a select expression appears outside of any sustainable block, the greatest mode is used. This aligns with our intuition that code not subject to sustainability management should not be approximated.

IV. THE ECO DESIGN: COMPILER SUPPORT

In this section, we describe compiler support, primarily answering Q3 and Q4.

A. Automated Calibration Insertion

We now define a compiler transformation to automatically identify and insert calib expressions. Several design choices are possible: (1) check periodically; (2) check whenever the supply gauge changes; (3) check whenever the demand gauge changes. Route (1) is ad hoc, especially when selecting a fixed “period” that fits all programs. For (2), the supply gauge in energy-aware programming is often associated with system-level variables (such as remaining battery) whose updates are often performed outside the program runtime and out of the
control of programmers. This may lead to platform-dependent behaviors: the same program may calibrate at different rates on two computers with different battery drivers.

Eco follows route (3). Demand gauge is typically an expression formed with program variables, such as loop index in the example in Figure 2. The state change of these variables is visible and predictable to the programmer. For application-level energy management strategies—a family that Eco belongs to—we believe the programmer should have control on program behaviors related to energy awareness.

We now define the transformation process. The transformation is isomorphic except that expression e is transformed to calib e iff \( \neg P_0(e) \land (P_1(e) \lor P_2(e)) \), where:

- \( P_0(e) \): e can only be reached from uniform blocks
- \( P_1(e) \): e is assignment, and the left-hand side variable appears in the demand gauge of lexically enclosing sustainable blocks
- \( P_2(e) \): e is a field write expression, and said field is read in the demand gauge of any sustainable block that may reach e

The implementation of the three predicates has little novelty. \( P_1 \) is a trivial local analysis. Our Eco implemented \( P_0 \) and \( P_2 \) through a standard context-sensitive reachability analysis. Furthermore, observe that when \( e \) is reachable from both uniform blocks and non-uniform blocks, calib will be inserted. This is sound because its evaluation will become a no-op when reached from a uniform block (\( \bot \) is in the S-stack).

B. Automated Mode Selection Insertion

The syntax we introduced earlier requires programmers to select mode cases explicitly through the select expression. For example, to assign a variable \( y \) of mcase(int) type to variable \( x \) of int type, programmers need to write \( x = \text{select } y \). This can be an onerous task for programmers, especially for incremental programming, where programmers start with a program written in an existing language, and incrementally "try out" new language abstractions. An Eco programmer may start with a Java program, and alter a variable from, say, holding a Res(1024, 768) object to the mode case we used in Figure 2. If explicit select is needed, the programmer would have to manually insert a select expression for every occurrence of such variables in a sustainable block where the mode case is meant to be selected.

We define a simple compiler transformation to relieve the programmers of the burden of explicitly declaring select expressions. The key insight is that Eco as a strongly typed language provides sufficient type information to compare both the sending and the receiving ends of a data flow. For example, if a programmer writes \( x = y \) in the program, and variable \( y \) is of mcase(int) type and variable \( x \) is of int type, our compiler transformation algorithm will transform it into \( x = \text{select } y \).

More generally, the transformation is hinged upon the definition of predicate \( \langle \tau \nearrow \tau' \rangle \# (e \nearrow e') \), which says an expression e of type \( \tau \) should be transformed to \( e' \) if it appears at a program point expecting an expression of type \( \tau' \). The predicate holds iff either of the two holds:

- \( \tau << \tau' \)
- \( \tau << \tau' \) does not hold, and \( \tau = \text{mcase}(\tau'') \), and \( (\tau'' \nearrow \tau') \# (e \nearrow e') \)

where the \(<\) relation is reflexive and transitive, with two additional rules. \( \tau << \tau' \) iff (i) \( \tau \) is a subclass of \( \tau' \); (ii) \( \tau = \text{mcase}(\tau_0) \) and \( \tau' = \text{mcase}(\tau'_0) \) and \( \tau_0 << \tau'_0 \).

Some examples should be sufficient to illustrate the ideas here:

\[
\begin{align*}
(\text{mcase(num)} &\nearrow \text{num}) \# (e \nearrow \text{select } e) \\
(\text{mcase(mcase(num))} &\nearrow \text{num}) \# (e \nearrow \text{select } (\text{select } e)) \\
(\text{mcase}(x) &\nearrow y) \# (e \nearrow \text{select } e) \quad \text{where } X \text{ is a subclass of } Y
\end{align*}
\]

Finally, the transformation is isomorphic, except that:

- **R1**: For every expression where (sub-)expression \( e \) flows to (sub-)expression \( e' \), where the type of \( e \) and \( e' \) are \( \tau \) and \( \tau' \) respectively, we transform the aforementioned expression to one the same as before, except that \( e \) is replaced with \( e'' \), and \( (\tau \nearrow \tau') \# (e \nearrow e'') \). Here, the "flows-to" relation is the standard notion in data flow analysis.

- **R2**: For every expression where the (sub-)expression \( e \) appears in a program point that serves as an object target (such as the receiver of the method invocation, or field read or write), we transform the aforementioned expression to one the same as before, except that \( e \) is replaced with \( e'' \), and \( (\tau \nearrow X) \# (e \nearrow e'') \), and \( \text{ATOM}(\tau') = X \).

where \( \text{ATOM}(\tau) \) is defined as \( \tau \), except that it is defined as \( \text{ATOM}(\tau') \) if \( \tau = \text{mcase}(\tau') \).

C. Type Checking

Type checking is defined over the post-translation programs where calib and select expressions are explicit. It is largely standard. The subtyping relation is defined to be identical as the \( <\) relation. For \( \{m_1 : e_1; \ldots; m_n : e_n\} \), we require the type of \( e_i \) for \( 1 \leq i \leq n \) to be a subtype of the declared type \( \tau \).

V. Programming Idioms

We now summarize common idioms we encountered while programming in Eco. Broadly, the summary may serve as a first step toward understanding design patterns and micro-patterns of sustainable software.

A. Abstract Supply Units

The sugared syntax \texttt{bsupply} use battery capacity (\texttt{mAh}) as the unit of supply, and syntax \texttt{tsupply} uses temperature in Celsius. These units are not hardcoded in Eco. More generally, an Eco programmer may choose any unit intuitive and accurate for her sustainability management, as long as such values can be obtained from underlying OS and hardware. For example, popular battery-powered consumer computers (laptops and smartphones) typically come with power management modules that can report the remaining operating time (or remaining percentage). A sustainable block that should consume no longer than 20% of the remaining operating time can be programmed as:
Battery-powered computer users often observe that the same program execution that reduces battery from 80% to 78%

```java
double budget = MAX;
old = battery;
sustainable {
    for (int num = 0; num < IMGS; num++) {
        if (num == 0.1*IMGS) {
            budget = (battery - old) * 9;
        }
    }
}
bsupply (budget)
demand (... ...) \Rightarrow (...) ...
```

where \texttt{getROT} queries the system battery management module for the remaining operating time.

In the real world, the suitable unit to report battery State-of-Charge (SoC) [14] is dependent on many factors, such as (i) the characteristics of batteries, e.g. lithium-based or nickel-based; (ii) the support of OS, e.g. ACPI-compliant systems and different vendor extensions (such as \texttt{BatteryService} in Android); (iii) the estimation algorithms. The more user-friendly metrics—remaining percentage, or remaining operating time—are in fact derived metrics from lower-level battery status data e.g., voltage level, current level, and coulomb counts [14]. Under the backdrop of this diverse and fast-changing landscape, the abstraction provided by \texttt{Eco} maintains a level of stability in the presence of technology changes.

### B. Fixed Supply Characterization

Various forms of fixed battery budget characterization can be directly supported:

- \textit{fixed absolute battery budget}, e.g., \texttt{bsupply(20000)} says the execution can consume up to 20000 mwh of battery.
- \textit{fixed relative battery budget}, e.g., \texttt{bsupply(0.2 * battery)} says the execution can consume up to 20% of the remaining battery.
- \textit{maximum battery budget}, e.g., \texttt{bsupply(battery)} says the execution can as much as the entire remaining battery.

Analogously, simple forms of temperature budget characterization are supported:

- \textit{fixed temperature threshold}, e.g., the use of \texttt{tsupply(70 – temperature)} says the execution should not increase the CPU temperature to more than 70°C.
- \textit{fixed temperature increase}, e.g., \texttt{tsupply(20)} says the execution should not increase the CPU temperature by more than 20°C.
- \textit{fixed temperature increase ratio}, e.g., \texttt{tsupply(1.1 * temperature)} says the execution should not increase the CPU temperature by more than 10%.

In \texttt{Eco}, both supply sum and supply gauge are expressions reevaluated every time a calibration is performed. This implies the \textit{budget} (the value of what the supply sum expression evaluates to) does not need to remain constant throughout the execution of the sustainable block. This feature can be useful to adaptively set the budget. For example, a programmer may not know at the beginning what a “fair” budget would be, and she can “test run” a portion of the program to find out and reset the supply budget. Consider the example in Figure 5. Here, the program decides to executes 10% of the images, and calculate the budget for the remaining 90% of the images processing accordingly.

More broadly, revisable budget programming pattern may also be useful in (1) energy-aware systems relying on renewable energy (e.g., solar), where supply change is not monotone as the sustainable block executes; (2) systems with adaptive cooling devices, where temperature change may also follow complex patterns.

### D. Adaptive Demand Characterization

Just like supply, demand sum and demand gauge are also expressions that are reevaluated upon calibration. This allows programmers to adaptively adjust the demand during the sustainable block execution. For instance, Figure 6 shows a refined version of Figure 2, where both the size of the overall demand (i.e., what the demand sum evaluates to) and the size of the remaining demand (i.e., what the demand gauge evaluates to) may be adjusted. Here, we not only consider the number of images as demand indicators, but also the number of pixels in the images. In the new scheme, the change of resolution would alter both the size of the overall demand and the size of the remaining demand, because the adoption of a lower resolution in rendering an image resulted in both the slower-than-expected growth for completed work (\texttt{workdone}), and the reduction of the projected work ahead (\texttt{workleft}).

### E. Non-Linear Behaviors

Battery-powered computer users often observe that the same program execution that reduces battery from 80% to 78%
may reduce the same battery from 5% to 1%. Such non-linear battery behaviors are generally minimized by built-in OS/firmware estimators for new batteries, but they may become more pronounced as batteries age. If high precision on battery supply is needed, such variations should not be ignored and can be modeled in Eco as:

1. sustainable {
   2.     double startB = 0.2 * getB();
   3.     ...;
   4. }
5. supply (startB) (startB - redress(getB()));
6. demand (...)

where redress is a custom-defined function mapping readings from getBP to the “compensated” values. For example, one simplistic redress function would be to map value $b$ to $b$ if $b \geq 0.2$ and $b$ to $0.9 \times b$ otherwise. As energy-aware programming becomes more prevalent, we envision such “redress” functions would be part of the “battery profile” available to end programmers as library APIs.

A similar programming idiom can be applied for non-linear thermal behaviors. According to solid matter physics, the energy to increase the temperature from 30°C to 35°C and that from 300°C to 305°C significantly differ. (The range of operating temperatures for CPUs is usually much narrower, so the non-linear effect in this area is relatively small.)

The same idea can be applied to demand characterization. Non-linear behaviors can be supported when redress is not a linear function.

VI. IMPLEMENTATION AND EVALUATION

A. Implementation

Eco is implemented on top of the Polyglot compiler framework 2.5 [28], as an extension to Java. The battery data was queried through Advanced Configuration and Power Interface (ACPI) [1] and read by the program. The CPU temperature was queried with the CoreTemp² API, in Celsius. CoreTemp collects data directly from a Digital Thermal Sensor (DTS) located in each individual processing core. There is no need for external measuring devices or external circuit located on the motherboard to report temperature.

B. Benchmarks

We selected the following Java benchmarks and modified them into Eco: (1) sunflow,³ a rendering system that uses raytracing, (2) jspider,⁴ a web crawler, (3) montecarlo,⁵ a financial simulation from the Java Grande benchmark suite, (4) xalan,⁶ an XSLT processor that converts XML documents, and (5) rasterizer, from the batik⁷ distribution, that rasterizes SVG files. Our main selection criterion is diversity, covering domains such as graphics, web, statistics, program transformation, and data transformation. These benchmarks cover a diverse range of applications to support variability in program behaviors, such as image resolution in sunflow, URL request intervals in jspider, sampling size in montecarlo, and rasterization quality in batik. Several selected benchmarks appeared in Dacapo® benchmark suite.

All experiments were performed on a Intel 2.53GHz Duo core CPU laptop with 8GB RAM.

C. Battery-Aware Programming

We modified the Java sunflow benchmark with Eco syntax. The rendering logic in sunflow resembles the code snippet in Figure 6. Our benchmarking execution renders 9 images (IMGS=9), and the resolution and depth are implemented as first-class mode cases with identical definitions as in Figure 2.

We first execute the unmodified Java program and use its battery consumption as the baseline supply. This number is used for setting the BUDGET constant in Figure 6. In our experiment, BUDGET = 20970mWh. We execute the Eco program with 90%, 70%, 50% of the BUDGET. It is worth pointing out that the experiment process described here is only meant for inducing “threshold behaviors”, e.g., the program running with “just enough” battery, or “slightly inadequate” battery, or “significantly inadequate” battery. An Eco programmer does not need to profile in this manner (or worse, plugging 20970 into the source code). The more intuitive programming idioms were described in Section V.

The time series of supply gauge (blue lines) and demand gauge (red lines) for the sunflow executions with 90%, 70%, 50% of the BUDGET were shown in Figure 7, Figure 8, Figure 9, respectively. All data are normalized. All experiments are started with the same battery level. At any given time, if the demand curve is (significantly) above the supply curve, the program runs in a “deficit.” A mode change will happen after the processing of the current task, in the hope that the demand for the next task will go below the supply change. As one can see, battery supply change is near linear. The change in demand happens at discrete time, since workdone and workleft are not updated until an image is processed.

²http://www.alcpu.com/CoreTemp
³http://sunflow.sourceforge.net/
⁴http://jspider.sourceforge.net/
⁵http://www.javagrande.org/
⁶http://xalan.apache.org/
⁷http://xmlgraphics.apache.org/batik/
⁸http://www.dacapobench.org/
Therefore, each “plateau” in the demand curve represents the processing of one image. As high-resolution rendering takes significantly longer time, the “width” of each “plateau” is indicative of the operational mode.

Sustainability—the scenario where the supply consistently meets the demand—is maintained in all three scenarios: observe the supply curve and the demand curve in all three figures follow similar patterns of decline. The following table reports the statistics of our experiments:

<table>
<thead>
<tr>
<th>PER</th>
<th>supply set</th>
<th>supply used</th>
<th># hi</th>
<th># mid</th>
<th># lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>188.6wh</td>
<td>186.0wh</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0.7</td>
<td>147.6wh</td>
<td>142.1wh</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>104.8wh</td>
<td>103.3wh</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Observe that in all cases, Eco runtime is capable of meeting the budget (i.e., the “supply used” is less than the “supply set”). The # hi, # mid, # lo columns in the table list the number of images rendered with high/mid/low resolutions. Predictably, more images are rendered with lower resolutions as the supply is set lower.

Next, we designed a battery-aware variant of batik, processing 27 SVG files with the supply of 8050mwh × 70%. Similar to the fashion the sunflow experiment was constructed, we first executed the unmodified Java program and obtained the battery usage of 8050mwh, which we informally call “the budget.” The battery-aware variant thus runs on 70% of that budget. The adaptive behaviors are defined by a mode case value for target image height. Since the image resolution is adjusted by ratio, this value is strongly correlated with the quality of rasterization, as a larger height means a larger width and therefore higher resolution. The mode case is defined by height pixel as \{ hi : 12000; mid : 10000; lo : 8000 \}, where 12000 is used for the original Java execution. As Figure 10 suggests, Eco achieves the goal of sustainability by matching supply and demand closely. Figure 11 shows the adaptively selected resolutions based on the changes of supply and demand.

Last, we modified montecarlo, where a mode case is defined to adjust the sample size. Specifically, the value in each case member is the difference between the “perfect” sample size (280,000) and the sample size being used for that mode. The larger sample size, the higher quality the output will be. We use a syntactic sugar to define this mode case (with a large number of cases whose mode names are implicit), in the same effect of defining \{ m_0 : 0;...;m_i : i × ∆;...;m_n : n × ∆ \} where \( ∆ = 10,000 \). In other words, our benchmark adjusts the sample size by the increment/decrement of \( ∆ \). The execution of the unmodified Java code (for Monte Carlo core algorithm) consumes 2060mwh of battery (“the budget”). The Eco execution operates on of 80% of that budget, with results shown in Figure 12 and Figure 13. As Figure 12 suggests, Eco again matches supply and demand closely. Figure 13 shows the change of sample size over time.

D. Temperature-Aware Programming

To demonstrate temperature-aware programming, we use xalan to transform 17 XML files with 8000 iterations. We modify the original Java program as follows. First, we set a fixed temperature threshold to 60°C. This is expressed as tsupply(60 − temperature). Second, in between every other file transformation, we allow the CPU to sleep at a fixed interval. A mode case is used for this sleep interval, set with two case members: when the mode is hi, the interval is set at 4 milliseconds, whereas when the mode is lo, the interval is set at 10 milliseconds.

Figure 14 demonstrates the temperature changes of xalan, both in the Java execution and the Eco execution. As a benchmark, xalan involves intensive CPU-bound computations. Without any sustainability management, the Java execution drives the temperature over 60°C within one minute (a rise of about 15°C). The Eco execution on the other hand successfully maintains the CPU temperature within the threshold. Figure 15 may explain the fundamental difference between the two executions, showing different levels of CPU utilization (Y-axis, plotted with OS performance monitor). During the Java execution, the CPU utilization is mostly 100%. The Eco execution however regulates the CPU utilization at a much lower adaptive level.

The figures here also demonstrate the trade-off between temperature regulation and performance. Without any sleep, the Java execution is able to complete the transformation within about half the time of the Eco execution (terminates after around 301 seconds). This comes with no surprise: the latter execution intentionally slows down itself frequently. Figure 16 demonstrates the change of sleep intervals over time.
We also constructed a temperature-aware execution of jspider in Eco. Similar to xalan, the temperature is also set at the fixed threshold of 60°C. We employ a similar strategy for temperature control, by allowing the crawler to sleep at fixed intervals. Unlike xalan where each XML file is processed in a relatively short period of time, each step in crawling—involving requesting a URL, parsing the webpage, and analyzing the links contained in the page—takes significantly longer. This allows us to design with longer intervals of sleeps (but much less frequent than xalan). The interval is implemented as a mode case, with hi mapping to 0.5 second, mid to 1 second, and hi to 2 seconds. Figure 17 shows the temperature results. Figure 18 shows the CPU utilization, and Figure 19 shows the interval change. Typical for a web crawler, there is no “completion” time unless the user stops the execution. Our demand is set (and reset) at fixed time intervals. As a result, all time series shown here for jspider can be viewed as the first 800 seconds of an (almost infinitely running) execution.

E. Programming Efforts

The programming effort for Eco is mild, and the LOCs that involve Eco-specific changes are relatively small:

<table>
<thead>
<tr>
<th>LOCs</th>
<th>sunflow</th>
<th>jspider</th>
<th>montecarlo</th>
<th>xalan</th>
<th>batik</th>
</tr>
</thead>
<tbody>
<tr>
<td>original</td>
<td>30984</td>
<td>33986</td>
<td>3128</td>
<td>347981</td>
<td>325238</td>
</tr>
<tr>
<td>Eco changes</td>
<td>46</td>
<td>56</td>
<td>23</td>
<td>32</td>
<td>27</td>
</tr>
</tbody>
</table>

Despite some benchmarks have high LOCs in their Java versions, the changes involved for sustainable programming concentrate on a very small number of files, typically the entry points of core algorithms, and the variables representing the adjustable parameters of those algorithms. In our experience, the majority of time was used in understanding existing benchmarks, and finding appropriate procedures/algorithms to support adaptive sustainable programming. Once the algorithm of interest is identified, the programming effort itself for developing the benchmarks we experimented with — adding Eco-specific expressions into programs, and debugging — is small, usually in hours.

VII. RELATED WORK

Sustainable computing is an emerging topic in computer science. Within software engineering, support for sustainability has appeared in requirements engineering [34], [29], [32] and software architecture [18], [20]. At lower levels of the compute stack, modeling computer systems with supply and demand is mostly known in operating systems. Examples include resource management and scheduling for grid environment [6], device energy management [40], utility computing in clouds [7], energy distribution among users [27]. Recently, the supply/demand model is also used for performance tuning [26]. Eco enriches the landscape of sustainable computing by empowering programmers.

A number of programming models exist for developing energy-aware software, such as Green [4], EnerJ [36], and ET [10]. They do not offer fine-grained energy/temperature budgeting to programmers, and do not focus on maintaining sustainability through programmer supply/demand characterization. Beyond energy-aware computing, programming models for self-adaptive systems (e.g., [8]) share a high-level design goal of Eco: promoting adaptiveness in the presence of dynamic variations of software/system environments.

Broadly, application-level energy management and optimization is an active area, with recent results diversely ranging from program analysis and optimization (e.g., [16], [5]), testing and debugging (e.g., [21], [2]), design patterns (e.g., [35], [24]), decision framework (e.g., [25]), measurement and estimation support (e.g., [37], [17]), runtime support (e.g., [33]), and empirical studies (e.g., [30], [31], [23]).

More broadly, there is a long history in VLSI, architecture, and OS research to design and implement battery-aware (e.g., [12]) or temperature-aware (e.g., [38]) computer systems. A fundamental challenge is the need to “predict the future,” with solutions ranging from profiling and monitoring [11], machine learning [9], to agent-based modeling [13]. Eco complements existing systems by bringing programmer knowledge into the design space.

VIII. CONCLUSION

This paper describes Eco, a sustainable programming model for developing Java-like applications. With novel abstractions for supply and demand shaping, Eco brings programmers into the loop of sustainability management and promotes energy-aware programming and temperature-aware programming.

Acknowledgments We thank the anonymous reviewers for their useful comments, and Anthony Canino for his help improving the compiler and the benchmarks. This work is partially supported by US NSF CCF-1054515.
REFERENCES

[34] Rohrer, K., and Richardson, D. Sustainability requirement patterns. In Requirements Patterns (RePa), 2013 IEEE Third International Workshop on (July 2013), pp. 8–11.