Evaluating Communication Pattern Representations in Execution Trace Gantt Charts - Supplemental Material

Category: n/a

1 Overview of Supplemental Materials

This document contains: additional figures and discussion about our experiments(Sect. 2), summaries of the four pilot studies we conducted (Sect. 3) and details regarding our design process(Sect. 4), including a table of visual channels and our rationale for using or rejecting them (Table 1) and images of rejected designs.

We also include a zipped archive containing the following:

1. A CSV including all codes from our preliminary study.
2. The design packet sent to our expert collaborators and the feedback they sent back.
3. The source code of our controlled user study.
4. Our analysis script and collected data from the main experiment and pilot experiments

2 Expanded Experiment Details

Here we include some figures and minor discussion which we could not include in the main paper due to space limitations.

2.1 Controlled User Study Tasks Figures

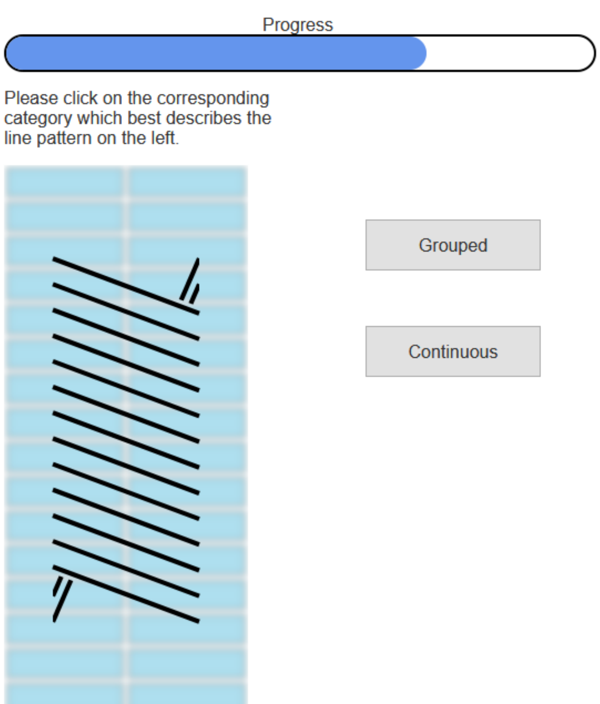

Fig. 1: This shows an example of a classification task where we asked participants to identify if a pattern was grouped or continuous.

Here we present multiple screenshots showing the tasks participants were asked to perform for our online experiment. Figure 1 and Figure 2 shows a examples of our classification questions where participants were asked to identify if a particular pattern and if that pattern was partitioned into groups or continuous. Figure 3 shows an example of an discrimination question where participants compared strides between two charts.

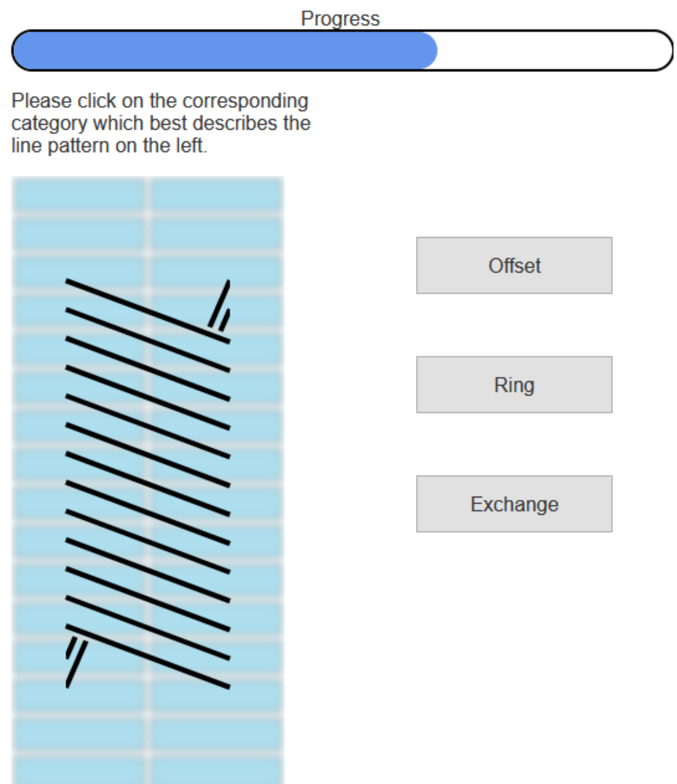

Fig. 2: This shows an example of a classification task where we asked participants to identify what pattern our chart exhibited.

We initially chose discrimination tasks based on the scenario that performance analysts might want to identify which regions in a Gantt chart are iterations of the same communication operation or even the same code. We switched pattern type and grouping to classification under the rationale that this identification was necessary to assess sameness and that recognition tasks for these higher level abstractions are important. We kept stride as a discrimination task because the exact number is less important and as stated previously, dependent on several execution-time calculations.

In addition to the tasks we show here, we also considered a drawing task, where participants would be asked to copy a pattern from one representation onto a blank background of a different size, as a holistic test. We ran a crayon and paper pilot of this approach but found the responses difficult to interpret, even with follow-up interviews.

2.2 Post Experiment Expert Feedback

For our post-experiment feedback we engaged in two 30 minute semi-structured interviews with experts familiar with Gantt charts in the

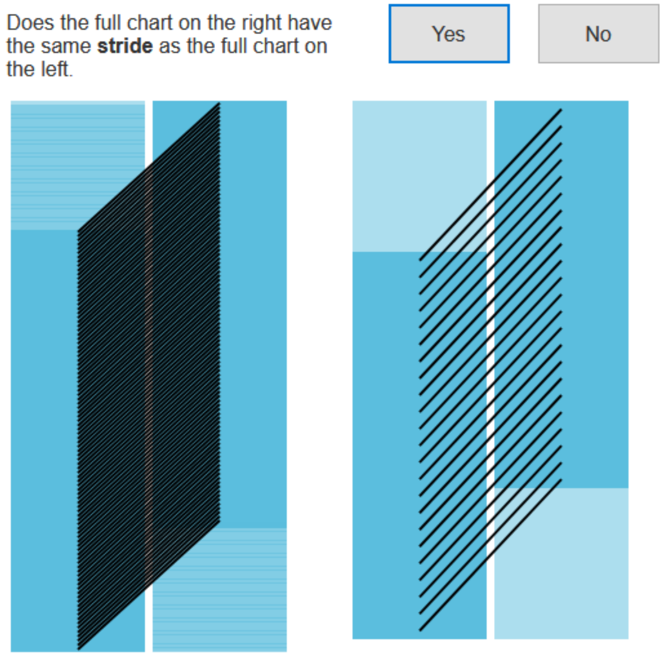

Fig. 3: This figure shows an example of a discrimination question. Users were asked to compare two charts and identify if they had the same stride or not.

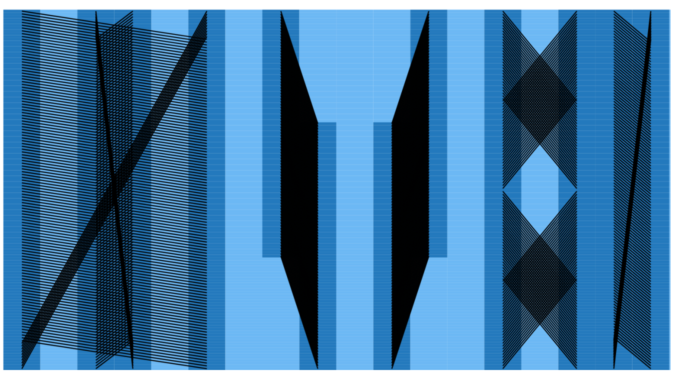

Fig. 4: A visualization mockup we showed to experts to elicit their feedback on the utility of our designs. This is a fully zoomed out example of a chart across many time steps.

HPC domain. One was performed online and the other in person. Each expert was introduced to the designs and encoding rules and shown a glyph-free chart at full (zoomed-out) (Figure 4) and partial (zoomed in) (Figure 5)representations and the same Gantt chart with glyphs overlaid following our speculative design rules (Figure 6). They were then asked questions we designed to understand the utility and weaknesses of our glyphs from an expert's perspective.

Specifically, they were asked:

1. Could you see something like this being useful?
2. What, if any, information could we change or add to make this more helpful to you?

The most salient responses and sentiments from these participants are included in the main paper.

3 Study Pilots

We summarize four pilot studies and how they influenced the design of the final study.

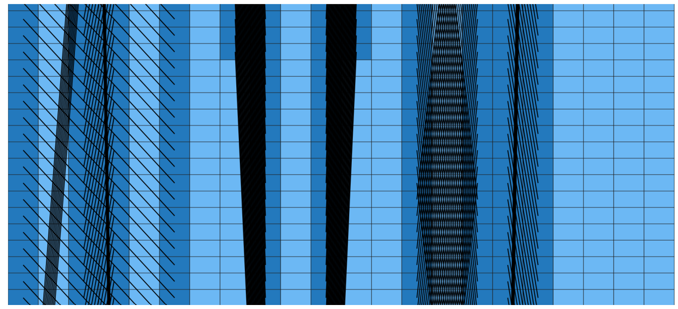

Fig. 5: A visualization mockup we showed to experts to elicit their feedback on the utility of our designs. This is a zoomed-in version of Figure 4

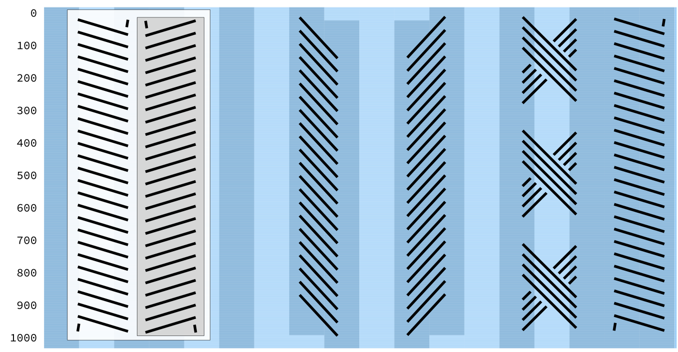

Fig. 6: A visualization mockup we showed to experts to elicit their feedback on the utility of our designs. Overlapping patterns are outlined with boxes and given a grey background.

3.1 First Offline Pilot: Multiple Choice

The first pilot considered only *full* and *partial* representations. The communication patterns used were offsets, rings, exchanges, and stencils.

After a brief introductory reading, participants were shown one target visualizations in one panel and four choice visualization. Each choice visualization had a checkbox. The participants were asked which off the choice visualizations were portraying the same communication pattern as the target visualization. The first three such questions were used as a tutorial. This task was similar to the discrimination task we used in our final study.

The pilot was conducted in-person on laptops provided by the authors. This allowed us to informally request feedback of the participants regarding the study as well as handle any unforeseen bugs in the software.

One rationale for testing a multiple-choice design with possibly multiple correct answers was to collect data about cases where one representation was recognized and one was not, with the ultimate goal of discerning what factors in the representation led to their efficacy. However, while participants seemed willing to choose multiple answers, they expressed confusion while doing so. Furthermore, we found it difficult to draw conclusions from the pilot data, given the multiple factors that could lead to an answer. Therefore, in future designs we decided to limit comparisons too a single pair.

3.2 Second Offline Pilot: Drawing

In addition to forced-choice tasks, we considered including drawing tasks, where participants would be asked to interpret a depiction of a pattern and then apply that pattern to a Gantt chart of a different scale. We designed a pen and paper pilot experiment to test this idea before proceeding with an electronic version. We recruited 15 participants at a regional hackathon.

We supplied participants with a packet of five questions, one per sheet, with each question containing either a *partial* or *full* rendering of communication lines over two columns of boxes simulating timestep-PE pairs in Gantt chart. For each question, next to the image, we printed an empty grid of boxes, two columns wide and 12 rows long. In both the *partial* and *full* cases, we asked users to draw a pattern over the 24 boxes equivalent to the pattern shown on the associated "chart." In addition to drawing lines, respondents were provided with crayons and asked to color in boxes to match the pattern of boxes on the left. Finally, participants were also asked to rate their perceived difficulty of each question and how confident they were in their correctness.

Before being given the study packet, individuals were introduced to the experiment with two basic questions of the same type as the rest of the experiment in a tutorial phase. During this phase, subjects were supervised by an experiment facilitator (author) and encouraged to ask questions as they completed this mini-packet. After the tutorial concluded, participants could no longer ask questions or seek assistance for the main packet of questions.

Responses were evaluated with a semi-subjective grading process where correctness was quantified on a 4 point scale. After this quantification it was determined that nearly half of all responses had to be thrown out due to fundamentally incorrect responses, an example of which can be seen in Fig. 7 and Fig. 8. In most of these cases, participants would replicate patterns very literally, disregarding the idea that one line connects two boxes and instead multiple lines spanning two boxes. These errors were like due to ambiguities in the question prompt and insufficient training.

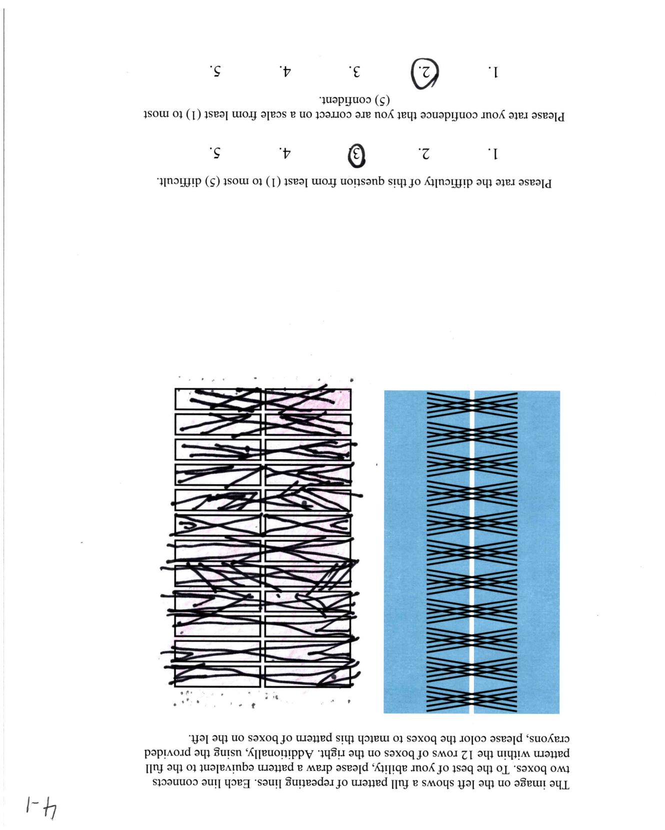

Fig. 7: An example of a fundamentally incorrect response from our "drawing" pilot where the participant drew multiple lines within and between boxes somewhat arbitrarily.

From the remaining data, we were unable to draw conclusions about how our independent variables of pattern and partial/full rendering type affected our dependant variable of response accuracy. Regardless, we did conclude that it would not be worth implementing this type of free-form experiment into a digital platform due with our current understanding of how people conceptualize patterns and thus how to train and screen for so open an experiment.

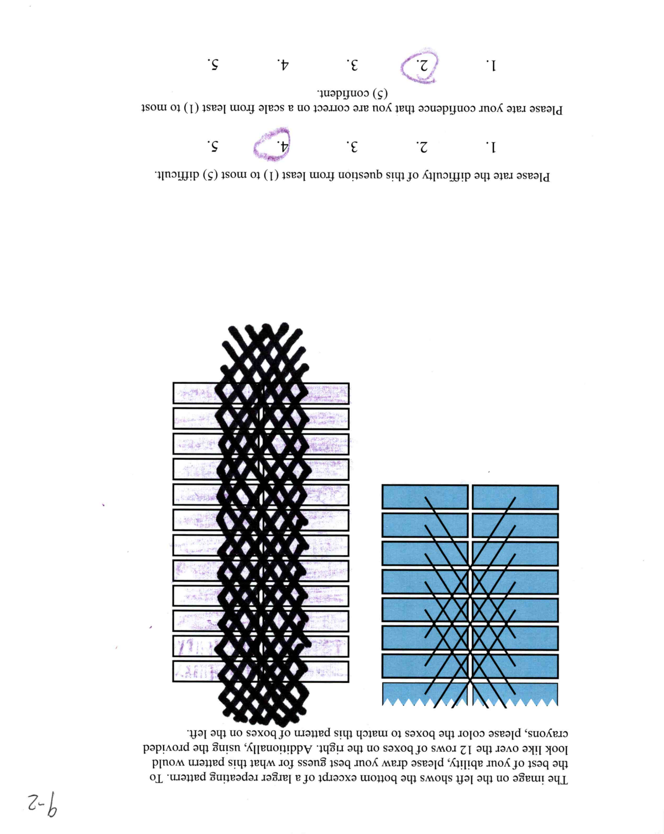

Fig. 8: A fundamentally incorrect answer where a participant connected lines which were not previously connected.

3.3 First Online Pilot

Our first online pilot was the first to include the new designs. We commissioned this pilot to 1) test the robustness of our experiment platform and 2) to further refine the experiment design.

This participants in this pilot were recruited from the research team's colleagues and friends. These participants provided feedback on the experiment itself, identified bugs as they came across them, and suggested quality of life improvements to the platform. The data from this particular pilot was analyzed but not closely scrutinized.

This pilot provoked several changes to the design of the experiment and experiment platform. First, the tutorial was substantially re-written to be less technical and several more example images were added. Second, we re-organized of the experiment into two sub-modules to address respondents comments that the tutorial was too long and they forgot details from early on by the time they got to the main questions. Third, comments from several participants about the uncertainty of if they were caught in a loop of questions lead to the introduction of the progress bar. Finally, we worked out minor bugs in the platform as they were identified.

3.4 Second Online Pilot

We ran our fourth and final pilot with goals: 1) testing the integration of the experiment platform with the Prolific platform, 2) providing data to develop the hypotheses we tested in our final experiment, 3) enabling a power analysis to estimate the desired number of participants for the final experiment, and 4) refining the time estimate given too participants for the full experiment.

This pilot had the same structure and question set as our final experiment. We put it out on the Prolific platform with a request for 10 participants. Ten responses were collected and finished the experiment within the hour of solicitation, all of which passed our guard questions.

The collected data was analyzed using the same techniques and analysis scripts as in our final experiment. As mentioned in the main paper, these results were approximately the same for *new* representations, but favored *full* representations compared to *partial*. This lead to our hypotheses which made comparative predictions about the performance of *partial* vs *full* representations. After performing a power analysis to arrive at the target of 35 participants, we leveraged our findings from this pilot to pre-register our experiment with OSF.

4 Design Process and Considerations

We describe our early design process of brainstorming and pencil and paper designs. This section is followed on subsequent pages by figures showing these early sketches as well as a table (Table 1) summarizing our rationale for the final visual channels used.

4.1 Brainstorming

We drew out many possible designs on paper with the intention of exploring the design space before committing to a particular design. The goal was to avoid premature design commitment. With the explicit goal of "exploring the space," we considered radical designs bordering on the absurd for this domain. In Fig. 9 and Fig. 10 there are several examples of unconventional designs, notably a heat-map applied to a Gantt chart showing the most dense overlap of communication, and ring timeline with arcing, edge bundled communication lines spanning processes at particular time steps. In addition to these more radical ideas, an early form of the designs implemented for this paper can be seen.

4.2 Pen and Paper Designs

After deciding to use metaphorical glyphs as our core design approach, we entered a phase of pen and paper design. Fig. 11 and Fig. 12 show early iterations of the proposed pattern representations highlighting variables and using channels which were later discarded. At this phase, we also see proposals for how to integrate these glyphs into a hypothetical Gantt chart. Ultimately a simpler design prevailed where we merely removed underlying communication data and show these glyphs permanently at dense zoom levels (a solution hinted at in Fig. 13). Another interesting note at this phase of design was the inclusion of color as a channel used to denote the beginning and ending of groups. Upon reflection this was ultimately deemed redundant and determined to be in violation of gestalt principles, so it was dropped.

Fig. 15 shows a pen and paper write up from a later design document where many encodings were considered but were ultimately abandoned. Color for showing grouping appears again to encode grouping in a different capacity, but was again dropped. Throughout these design phases it can be seen that we considered encoding the number of communication lines in a pattern, but this variable was deemed not especially salient at this phase as the structure of a pattern is far more informative than the "length" which is related to the number of processing elements anyway. Continuity, as described here, was included in the final designs of our glyphs for intermediate zoom levels which may still be sufficiently dense. It was also implemented for the experiment platform but was not evaluated as it did not apply to our larger research questions.

Table 1 summarizes the channels we considered in our design and our rationale regarding whether to use them and how.

Position on a Common Scale	We will likely place the entire glyph/icon on a common scale to convey that different patterns are associated with different columns. The common scale will be the regular time intervals denoted on the x axis.
Position on an Unaligned Scale	The unaligned scale does not seem to meaningfully convey time of a glyph in the same way that an common scale does. However, we can use it to convey grouping by aligning duplicates of glyphs on-top of one another in the y-axis. We could also place the entire glyph down on an unaligned scale as well. It will show that two glyphs are distinct. This could only likely work for the x axis.
Length	We cannot directly encode line length as it is directly related to the distance between calls to locations and the time step between calls as well. Line length is somewhat forced by timestep/angle so we may not be able to use it to directly encode. We don't understand how it may be used in that case but it might be a tradeoff
Tilt/Angle	Angle could be used to encode or convey or hint at the offset of our pattern. However, we have determined that offset is not a salient characteristic under many circumstances and will only be relevant when crossing nodes.
Area	Area could be used to convey that a pattern applies to a pair of processes at logical timesteps. It would encode that the patterns are referring to the horizontal and vertical space between t and t-1.
Depth	We do not think depth will be a very helpful encoding attribute at this time. For our 2d focused designs it would be expected to complicate the design without adding very much.
Luminance/Saturation/Opacity	Opacity could be used to convey boolean data of continuation off the edge of the drawing area. We can duplicate the patterns we have already drawn and render at a lower opacity them above or below our pattern, where they continue.
Curvature	Curvature is not needed to convey any information in our dataset. At our current iteration we are using only straight lines and rectangular boxes.
Volume	Our current data is plotted along two axes right now and we do not expect that encoding some value in 3D will convey information which he have not conveyed elsewhere.
Spatial Region	Spatial region could be used to convey the scope of a certain pattern in a way similar to area.
Color Hue	Color could be used for conveying grouping, possibly by showing boundaries between when one group starts and another ends. This will likely violate gestalt principles of similarity however.
Motion	Motion will likely not be used because it is a very strong popout channel and will be too distracting.
Shape	Shape will encode the base structure of a pattern. Even though it will be a shape made up of repeating-angled lines, it's still a shape.
Closure	We may enclose each of our pattern glyphs in a box or icon to draw the attention of users to that portion of the chart.
Density/Contrast	Density could be used to convey (very roughly) how many lines are contained within a group. This does pose problems of making the pattern unrecognizable with groupings of many-many lines.
Number	Like density, the number of parallel lines in a particular pattern could convey or hint at the total number of lines the underlying pattern has. Perhaps with a mapping of 1 line = 1000 real lines. Number could be used to convey grouping by doubling up a base pattern.
Orientation	Since angle will be used to convey to the user what pattern our software is encoding, mapping orientation to some variables seems like it would only make the visualization more confusing.

Table 1: A table of possible channels considered and thoughts on accepting/ rejecting them

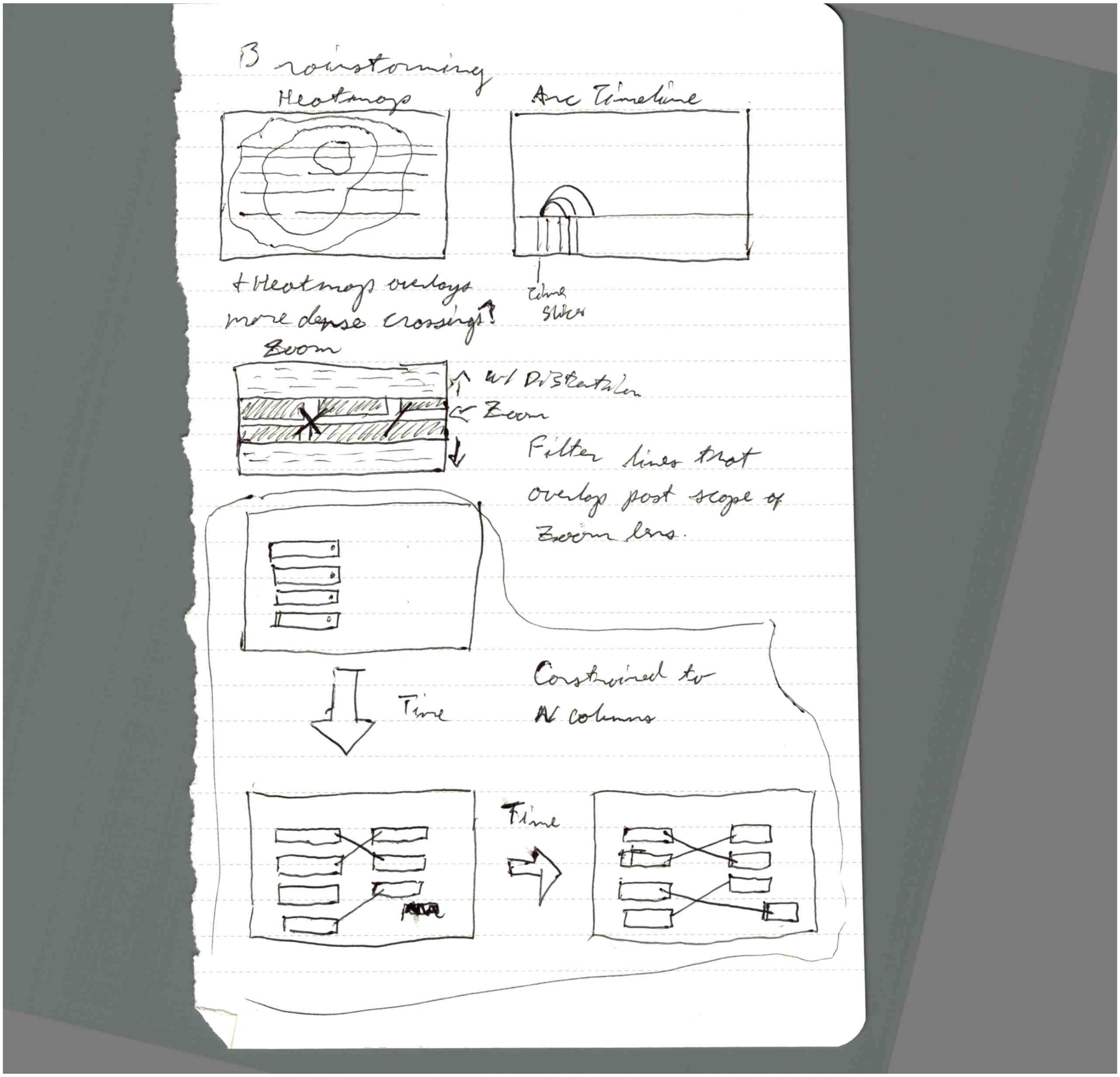

Fig. 9: Brainstorming document of possible early designs for a scale agnostic representation of communication patterns in Gantt charts.

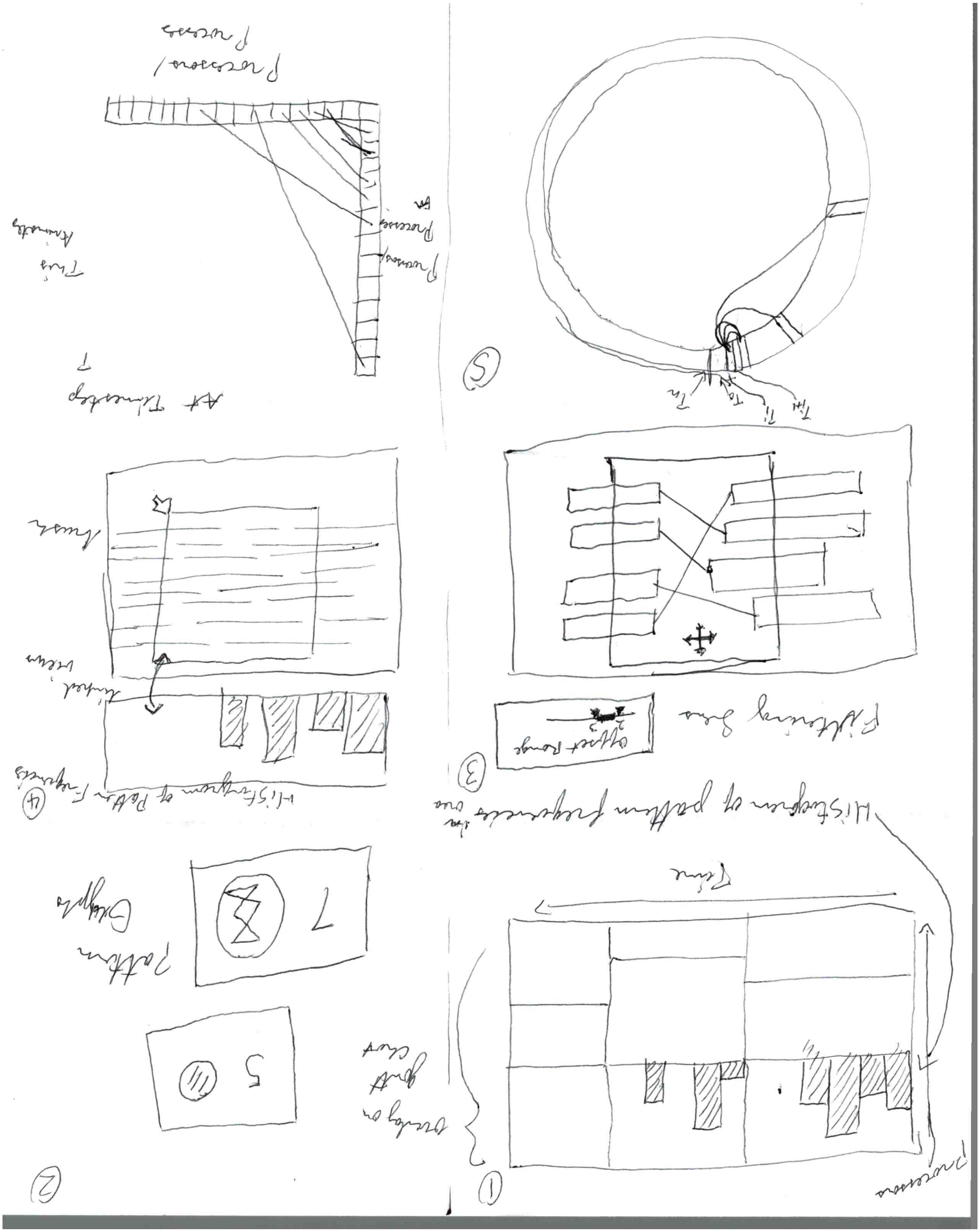

Fig. 10: Brainstorming document with many proposed designs for a scale agnostic representation of communication patterns in Gantt charts.

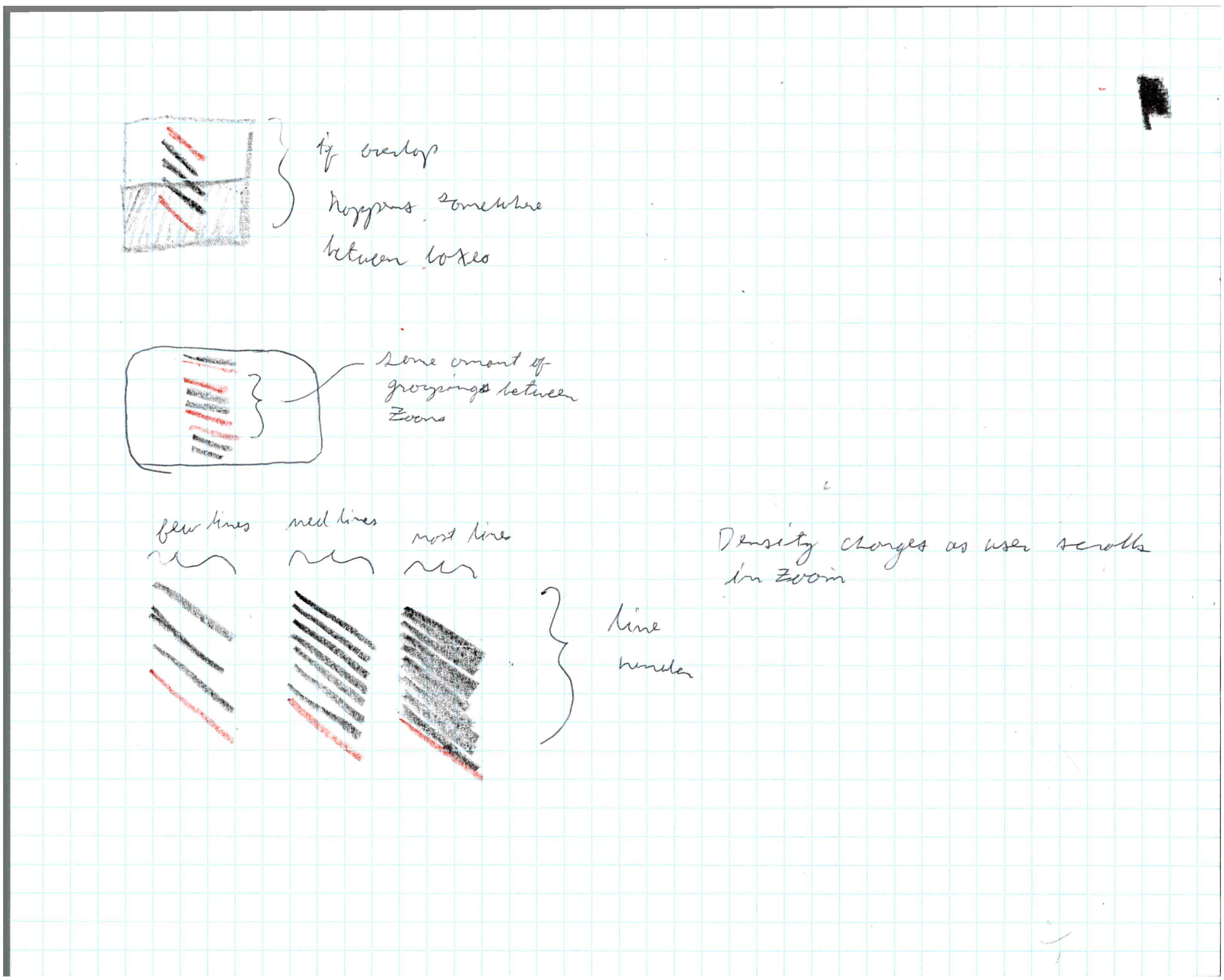

Fig. 11: An pen and paper design document detailing possible methods of rendering glyphs in a Gantt chart.

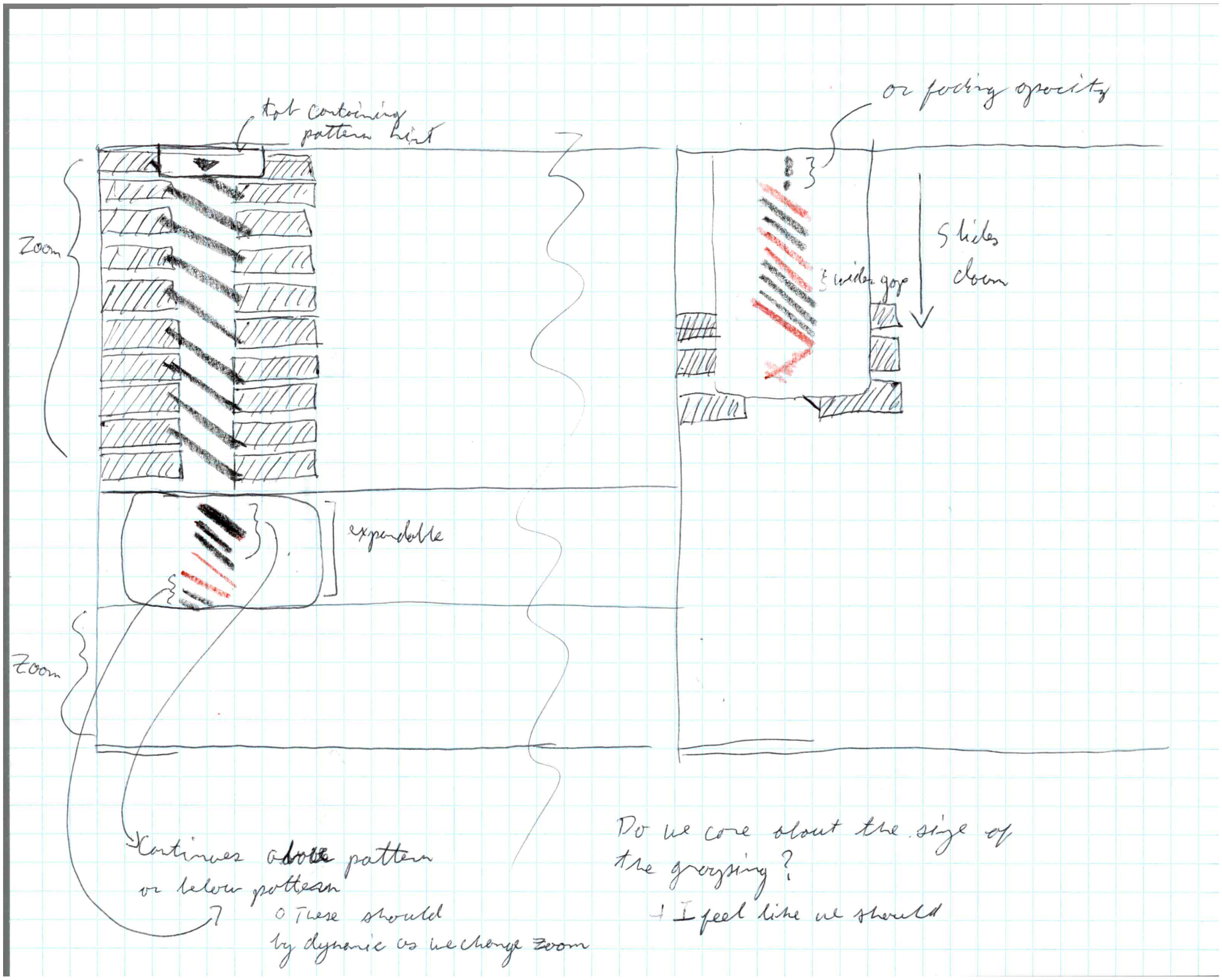

Fig. 12: A pen and paper design document detailing some potential encoding rules for proposed glyphs. We see here a proposed rule for when communication crosses nodes and an means to convey zoom levels with these patterns.

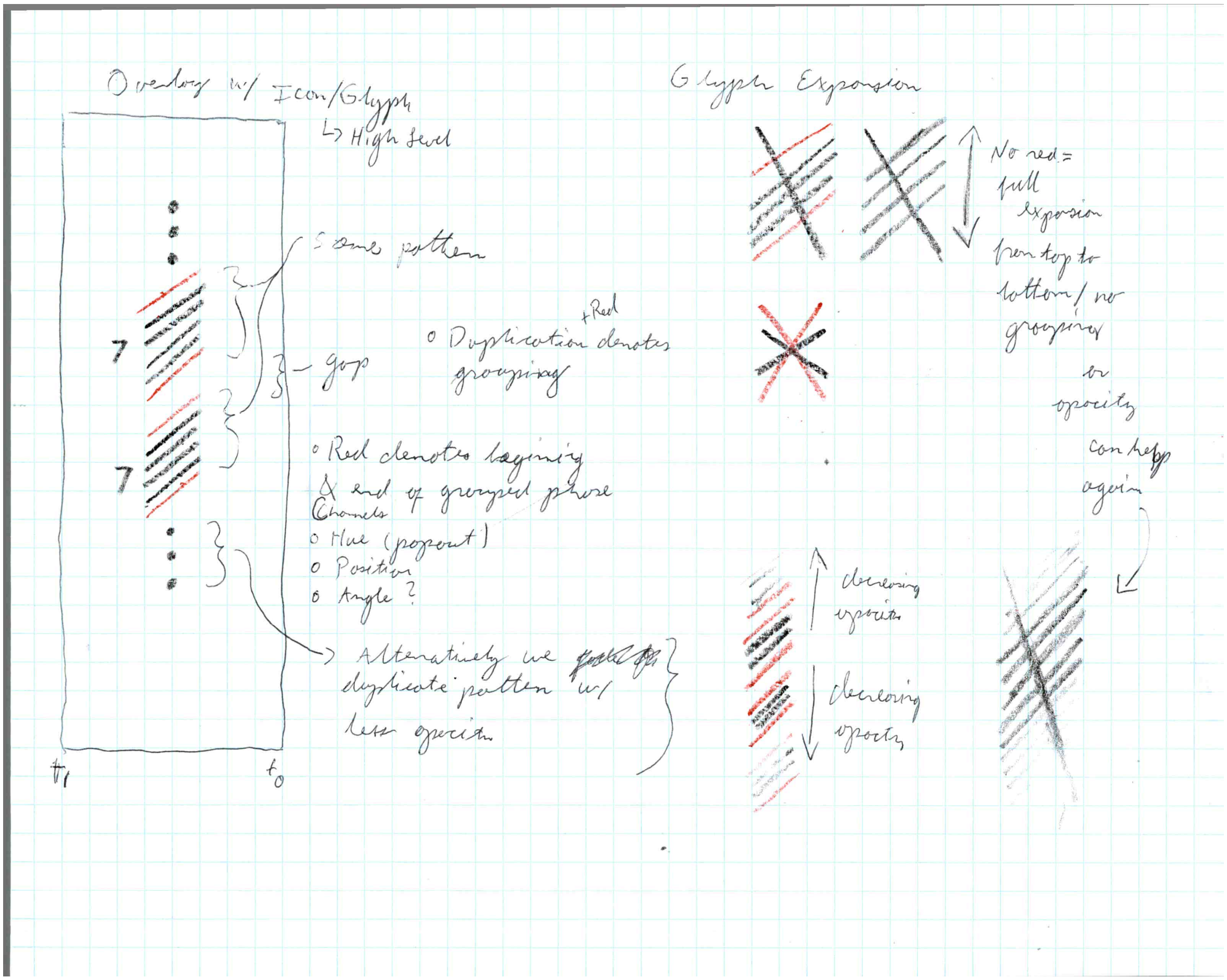

Fig. 13: A design document detailing some possible encoding rules and showing another way that these glyphs could be rendered on a Gantt chart. This document proposes many ways of conveying continuation of a pattern beyond the top or bottom of the drawn canvas.

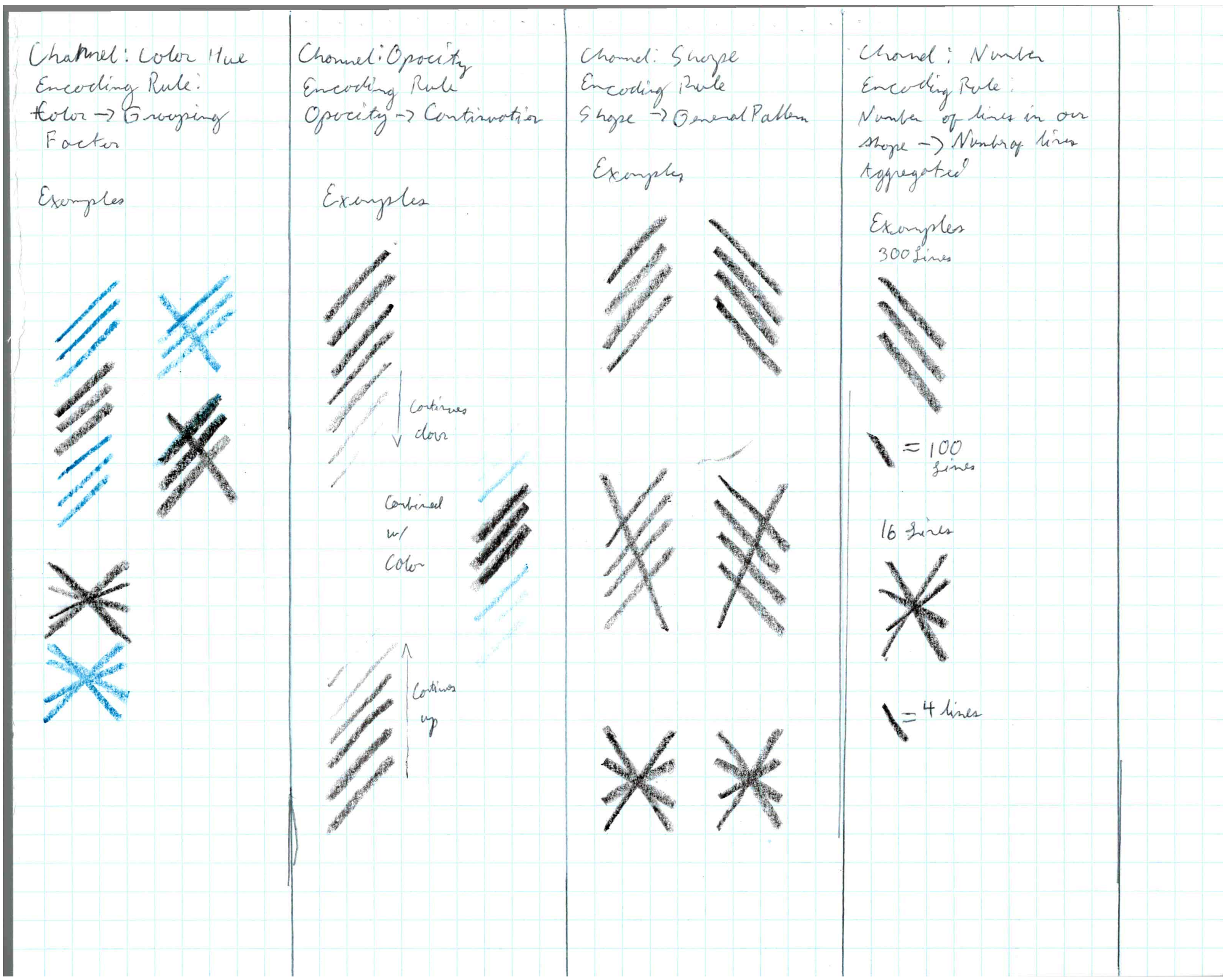

Fig. 14: A more formal design document explicitly stating proposed channels and encoding rules.

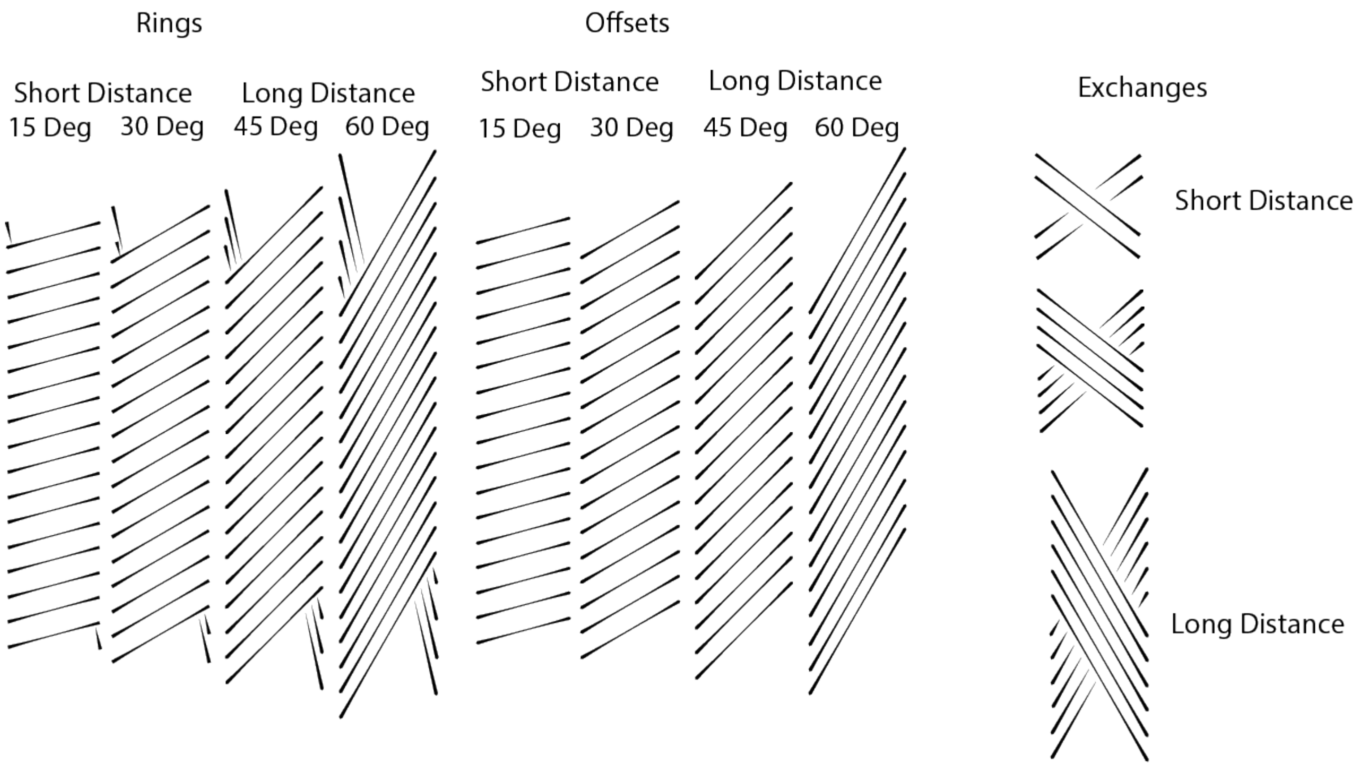

Fig. 15: A late stage design document showing what a visually refined set of glyphs will look like and how angle, encoding stride, will affect the look of each particular shape. The flared "pen" style applied to individual lines was added to convey that the lines in the glyph are abstractions and not the same as the underlying lines making up a communication pattern.